Springer Proceedings in Mathematics & Statistics

Sergei Silvestrov
Anatoliy Malyarenko
Milica Rančić *Editors*

Stochastic Processes and Applications

SPAS2017, Västerås and Stockholm, Sweden, October 4–6, 2017

Springer

Springer Proceedings in Mathematics & Statistics

Volume 271

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Sergei Silvestrov · Anatoliy Malyarenko
Milica Rančić
Editors

Stochastic Processes and Applications

SPAS2017, Västerås and Stockholm, Sweden, October 4–6, 2017

Springer

Editors
Sergei Silvestrov
Division of Applied Mathematics, School of Education, Culture and Communication
Mälardalen University
Västerås, Sweden

Milica Rančić
Division of Applied Mathematics, School of Education, Culture and Communication
Mälardalen University
Västerås, Sweden

Anatoliy Malyarenko
Division of Applied Mathematics, School of Education, Culture and Communication
Mälardalen University
Västerås, Sweden

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-030-02824-4 ISBN 978-3-030-02825-1 (eBook)
https://doi.org/10.1007/978-3-030-02825-1

Library of Congress Control Number: 2018958926

Mathematics Subject Classification (2010): 00A69, 60-XX

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedicated to
Professor Dmitrii S. Silvestrov's
70th birthday

Preface

This book highlights the latest advances in stochastic processes, probability theory, mathematical statistics, engineering mathematics and applications of algebraic structures, with a focus on those mathematical models, structures, concepts, problems and computational methods and algorithms that are important for applications in modern technology, engineering and the natural sciences. In particular, the book features mathematical methods and models from probability theory, stochastic processes, applied algebraic structures and computational modelling with various applications.

The book gathers selected, high-quality contributed chapters from several large research communities working on modern stochastic processes, algebraic structures and their interplay and applications. The chapters cover both theory and applications, and are illustrated with a wealth of figures, schemes, algorithms, tables and findings to help readers grasp the material, and to encourage them to develop new mathematical methods and concepts in their future research. Presenting new methods and results, reviews of cutting-edge research, and open problems and directions for future research, they will serve as a source of inspiration for a broad range of researchers and research students in, e.g. probability theory and mathematical statistics, applied algebraic structures, and applied mathematics.

This work arose on the basis of contributions presented at the International Conference "Stochastic Processes and Algebraic Structures — From Theory Towards Applications" (SPAS2017), which was held in honour of Professor Dmitrii Silvestrov's 70th birthday and his 50 years of fruitful service to mathematics, education and international cooperation. This international conference brought together a selected group of mathematicians, researchers from related subjects and practitioners from industry who actively contribute to the theory and applications of stochastic processes and algebraic structures, methods and models. It was co-organised by the Division of Applied Mathematics, Mälardalen University, Västerås and the Department of Mathematics, Stockholm University, Stockholm and held in Västerås and Stockholm, Sweden on 4–6 October 2017.

Representing the first of two volumes, the book consists of 19 chapters (papers), starting with a special chapter devoted to biographical notes about Professor Dmitrii Silvestrov and written by Sergei Silvestrov, Ola Hössjer, Anatoliy Malyarenko and Yuliya Mishura. The remaining 18 chapters are grouped into Part I — Stochastic Processes, and Part II — Applications of Stochastic Processes.

Part I begins with Chap. 2 by Dmitrii Silvestrov, which presents a survey of research results obtained by him and his colleagues in the areas of limit theorems for Markov-type processes and randomly stopped stochastic processes, renewal theory and ergodic theorems for perturbed stochastic processes, quasi-stationary distributions for perturbed stochastic systems, methods of stochastic approximation for price processes, asymptotic expansions for nonlinearly perturbed semi-Markov processes, and applications of the above results to queuing systems, reliability models, stochastic networks, bio-stochastic systems, perturbed risk processes and American-type options.

Chapter 3 by Dmitrii Silvestrov presents results of the complete analysis and classification of individual ergodic theorems for perturbed alternating regenerative processes with semi-Markov modulation. New short, long and super-long time ergodic theorems for regularly and singularly perturbed alternating regenerative processes are provided.

Chapter 4 by Sergey Krasnitskiy and Oleksandr Kurchenko explores asymptotics for Baxter-type sums for generalised random Gaussian fields. The general results are illustrated on examples related to generalised fields with independent values and the field of fractional Brownian motion.

In Chap. 5 by Salwa Bajja, Khalifa Es-Sebaiy and Lauri Viitasaari, upper bounds for rates of convergence in limit theorems for quadratic variations of the Lei–Nualart process are presented.

Chapter 6 by Yuliya Mishura, Kostiantyn Ralchenko and Sergiy Shklyar focuses on parameter estimation in the regression Gaussian model with discrete and continuous time observations. General results are applied to a number of models such as fractional Brownian motion, mixed fractional Brownian motion and sub-fractional Brownian motion, as well as a model with two independent fractional Brownian motions.

In Chap. 7 by Gulnoza Rakhimova, new effective conditions of asymptotic consistency for fixed-width confidence interval estimators and asymptotic efficiency of stopping times are highlighted.

In Chap. 8 by José Igor Morlanes and Andriy Andreev, an effective algorithm for the simulation of fractional Ornstein–Uhlenbeck process of the second kind is developed.

Chapter 9 by Kristoffer Lindensjö focuses on an extension of the constructive martingale representation theorem, from the space of square integrable martingales to the space of local martingales.

In Chap. 10 by Anatoliy Malyarenko and Martin Ostoja-Starzewski, random fields related to the symmetry classes of second-order symmetric tensors are described.

Part II begins with Chap. 11 by Dmitrii Silvestrov, Mikael Petersson and Ola Hössjer, in which asymptotic expansions for stationary and conditional quasi-stationary distributions of nonlinearly perturbed birth–death-type semi-Markov models are presented, and their applications to models of population growth, epidemic spread of disease and dynamics of the genetic composition of a given population are discussed.

Chapter 12 by Ola Hössjer, Günter Bechly and Ann Gauger addresses waiting times for coordinated mutations in a population described by Markov process of Moran type. The authors conduct a detailed analysis of the corresponding forms and conditions for waiting time asymptotics.

Chapter 13 by Kristoffer Spricer and Pieter Trapman presents the results of experimental studies of the initial phase of epidemic growth in large mostly susceptible populations. The authors demonstrate that the empirical networks tested in their paper display exponential growth in the early stages of the epidemic, except in cases where the networks are restricted by strong low-dimensional spatial constraints.

Chapter 14 by Elena Boguslavskaya, Yuliya Mishura and Georgiy Shevchenko studies Wiener-transformable markets, where the driving process is provided by an adapted transformation of a Wiener process. The authors also investigate the conditions of replication contingent claims on such markets.

Chapter 15 by Guglielmo D'Amico, Fulvio Gismondi and Filippo Petroni investigates the high frequency dynamic of financial volumes of traded stocks using a semi-Markov model. The authors show that this model can successfully reproduce several empirical facts about volume evolution like time series dependence, intra-daily periodicity and volume asymmetry.

In Chap. 16 by Benard Abola, Pitos Seleka Biganda, Christopher Engström, John Magero Mango, Godwin Kakuba and Sergei Silvestrov, new recurrent algorithms with linear time complexity for computing PageRanks for information networks with different graph structures are described.

In Chap. 17 by Pitos Seleka Biganda, Benard Abola, Christopher Engström, John Magero Mango, Godwin Kakuba and Sergei Silvestrov, traditional PageRanks based on an ordinary random walk approach and Lazy PageRanks based on a lazy random walk on a graph interpretation are considered. Further, the paper describes how the two variants change when complete graphs are connected to a line of nodes, the links of which are all in one direction. Explicit formulas and numerical results are obtained for both PageRank variants.

Chapter 18 by Hannes Malmberg and Ola Hössjer examines continuous approximations of discrete choice models with a large number of options. The authors use point process theory and extreme value theory to derive analytic expressions for the continuous approximations under a wide range of distributional assumptions.

Chapter 19 by Boris Faybishenko, Fred Molz and Deborah Agarwal presents the results of sensitivity analysis for nonlinear dynamics simulation of ecological processes based on models of deterministic chaos and their comprehensive time series analysis in the time domain and phase space.

This book was made possible by the strategic support offered by Mälardalen University to the research environment Mathematics and Applied Mathematics (MAM) in the established research specialisation of Educational Sciences and Mathematics at the School of Education, Culture and Communication at Mälardalen University. We are grateful to both the Department of Mathematics, Stockholm University and to the Mathematics and Applied Mathematics research environment MAM, Division of Applied Mathematics, School of Education, Culture and Communication at Mälardalen University, Västerås for their valued support and cooperation in jointly organising the successful international conference SPAS2017, which led to this book. We also wish to extend our thanks to the Swedish International Development Cooperation Agency (Sida) and International Science Programme in Mathematical Sciences (ISP), the Nordic Council of Ministers Nordplus Programme, and many other national and international funding organisations, as well as the research and education environments and institutions of the individual researchers and research teams who were instrumental to the success of SPAS2017 and to this book. In closing, we wish to especially thank all of the authors for their excellent contributions to this book. We also wish to thank the staff of the publisher, Springer, for their outstanding support with this book. All of the chapters have been reviewed, and we are grateful to the reviewers for their diligence.

Västerås, Sweden
July 2018

Sergei Silvestrov
Anatoliy Malyarenko
Milica Rančić

Contents

Contributors

Benard Abola Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Västerås, Sweden; Department of Mathematics, School of Physical Sciences, Makerere University, Kampala, Uganda

Deborah Agarwal Lawrence Berkeley National Laboratory, Computer Research Division, University of California, Berkeley, CA, USA

Andriy Andreev Department of Statistics, Stockholm University, Stockholm, Sweden

Salwa Bajja National School of Applied Sciences—Marrakesh, Cadi Ayyad University, Marrakesh, Morocco

Günter Bechly Biologic Institute, Redmond, WA, USA

Pitos Seleka Biganda Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Västerås, Sweden; Department of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam, Dar es Salaam, Tanzania

Elena Boguslavskaya Department of Mathematics, Brunel University London, Uxbridge, UK

Guglielmo D'Amico Department of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Chieti, Italy

Christopher Engström Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Västerås, Sweden

Khalifa Es-Sebaiy Department of Mathematics, Faculty of Science, Kuwait University, Kuwait, Kuwait

Boris Faybishenko Lawrence Berkeley National Laboratory, Energy Geosciences Division, University of California, Berkeley, CA, USA

Ann Gauger Biologic Institute, Redmond, WA, USA

Fulvio Gismondi Department of Economic and Business Science, University "Guglielmo Marconi", Rome, Italy

Ola Hössjer Department of Mathematics, Stockholm University, Stockholm, Sweden

Godwin Kakuba Department of Mathematics, School of Physical Sciences, Makerere University, Kampala, Uganda

Sergey Krasnitskiy Kyiv National University of Technology and Design, Kyiv, Ukraine

Oleksandr Kurchenko Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Kristoffer Lindensjö Department of Mathematics, Stockholm University, Stockholm, Sweden

Hannes Malmberg Stanford Institute for Economic Policy Research, Stanford, CA, USA; Department of Economics, University of Minnesota, Minneapolis, MN, USA

Anatoliy Malyarenko Division of Applied Mathematics, School of Education Culture and Communication, Mälardalen University, Västerås, Sweden

John Magero Mango Department of Mathematics, School of Physical Sciences, Makerere University, Kampala, Uganda

Yuliya Mishura Department of Probability Theory, Statistics and Actuarial Mathematics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Fred Molz Environmental Engineering and Earth Sciences Dept, Clemson University, Anderson, SC, USA

José Igor Morlanes Department of Statistics, Stockholm University, Stockholm, Sweden

Martin Ostoja-Starzewski University of Illinois at Urbana-Champaign, Urbana, IL, USA

Mikael Petersson Statistics Sweden, Stockholm, Sweden

Filippo Petroni Department of Economy and Business, University of Cagliari, Cagliari, Italy

Gulnoza Rakhimova Tashkent Auto-Road Institute, Tashkent, Uzbekistan

Kostiantyn Ralchenko Department of Probability Theory, Statistics and Actuarial Mathematics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Georgiy Shevchenko Department of Probability Theory, Statistics and Actuarial Mathematics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Sergiy Shklyar Department of Probability Theory, Statistics and Actuarial Mathematics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Dmitrii Silvestrov Department of Mathematics, Stockholm University, Stockholm, Sweden

Sergei Silvestrov Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Västerås, Sweden

Kristoffer Spricer Department of Mathematics, Stockholm University, Stockholm, Sweden

Pieter Trapman Department of Mathematics, Stockholm University, Stockholm, Sweden

Lauri Viitasaari Department of Mathematics and System Analysis, Aalto University School of Science, Aalto, Finland

Chapter 1
Dmitrii S. Silvestrov

Sergei Silvestrov, Ola Hössjer, Anatoliy Malyarenko and Yuliya Mishura

Abstract This chapter presents short biographical notes about Professor Dmitri S. Silvestrov.

Keywords Kiev University · Umeå University · Luleå Technical University Mälardalen University · Stockholm University

S. Silvestrov (✉) · A. Malyarenko
Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Box 883, 72123 Västerås, Sweden
e-mail: sergei.silvestrov@mdh.se

A. Malyarenko
e-mail: anatoliy.malyarenko@mdh.se

O. Hössjer
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
e-mail: ola@math.su.se

Y. Mishura
Department of Probability Theory, Statistics and Actuarial Mathematics, Taras Shevchenko National University of Kyiv, 64 Volodymyrska, Kyiv 01601, Ukraine
e-mail: myus@univ.kiev.ua

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_1

Dmitrii S. Silvestrov was born in 1947. D. Silvestrov graduated with distinction from Kiev University (Faculty of Mechanics and Mathematics) in 1968 and became a postgraduate student at the Department of Theory of Probability and Mathematical Statistics, under supervision of Professor M. Yadrenko. In 1969, D. Silvestrov defended the Candidate of Science (Ph.D. equivalent) dissertation [1], devoted to limit theorems for semi-Markov processes. In 1973, D. Silvestrov has got the Doctor of Science degree in the area of theory of probability and mathematical statistics. In the second dissertation, [2], D. Silvestrov developed an advanced theory of limit theorems for randomly stopped stochastic processes, which gives effective general conditions for weak convergence and convergence in topologies U and J for compositions of càdlàg processes. The above research directions were formed in the frame of the internationally known Ukrainian school on stochastic processes led by Academicians V. Korolyuk and A. Skorokhod.

The research results obtained by D. Silvestrov during this period are presented in the book [3]. In 1973, D. Silvestrov was awarded the Prize of the Moscow Mathematical Society on the recommendation of Academician A. Kolmogorov, who was at that time the President of this society, and, in 1977, the Ukrainian Ostrovsky Prize, for works on stochastic processes. The extended variant of the theory of limit theorems for randomly stopped stochastic processes is presented in the book [9].

In 1974, D. Silvestrov has got a Professor position at the Department of Theory of Probability and Mathematical Statistics at Kiev University. In the end of 70th, the research interests of D. Silvestrov shifted to the renewal theory and ergodic theorems for perturbed stochastic processes. The main results of this period are connected with a generalisation of the classical renewal theorem to the model of perturbed renewal equations and the exact coupling and ergodic theorems for perturbed regenerative and semi-Markov type processes. Partly, the results in this area are presented in the book [5]. Also, the book [4], co-authored with Professors A. Dorogovtsev, A. Skorokhod and M. Yadrenko, and containing the extended collection of probability problems, was published during these years.

In 80th, D. Silvestrov was also involved in an applied statistical research in cooperation with industry and development of statistical software. His interests in this area are reflected in the two books [6, 7].

One of his projects in the area of applied statistics connected with the database of statistical terminology attracted interest of Professor G. Kulldorff, who was at that time the President of International Statistical Institute. In 1991, he invited D. Silvestrov to continue this work at Umeå University. The comprehensive Elsevier's dictionary of statistical terminology [8], co-authored with Dr. E. Silvestrova, was published in 1995. In this way, D. Silvestrov began and then continued his work in Sweden, at Umeå University, Luleå Technical University, Mälardalen University, and Stockholm University.

During this period, D. Silvestrov developed a wide international scientific cooperation. Among his collaborators and co-authors are Professors M. Gyllenberg (University of Helsinki), J. Teugels (Catholic University of Leuven), R. Manca (University of Rome "La Sapienza"), Yu. Mishura and A. Kukush (Kiev University), G. Kulldorff, B. Ranneby, and H. Wallin (Umeå University), O. Hössjer and

A. Martin-Löf (Stockholm University), S. Silvestrov and A. Malyarenko (Mälardalen University). D. Silvestrov also was a visiting Professor at the Hebrew University of Jerusalem, the University of Turku, and the University of Rome "La Sapienza"; took part in organisation of a number international conferences and delivered invited lectures at these and many other conferences.

The cooperation with Ukrainian colleagues was continued in the frame of four European Tempus projects, which have been coordinated by D. Silvestrov and promoted creation and development of a specialty Statistics at Kiev and Uzhhorod universities, creation of a training center for actuaries and financial analysts at Kiev University and opening of a new specialty Educational Measurements at several Ukrainian universities as well as co-organization of three Scandinavian-Ukrainian conferences in mathematical statistics and 11 international summer schools in financial and insurance mathematics, educational measurements and related fields held in Ukraine and Sweden. D. Silvestrov is also a long-term member in the Editorial Boards of international journals Theory of Probability and Mathematical Statistics and Theory of Stochastic Processes.

In the beginning of 90th, D. Silvestrov began research studies of quasi-stationary phenomena in perturbed stochastic systems with random lifetimes. The results of these research studies are presented in the comprehensive book [10], co-authored with Professor M. Gyllenberg.

In 1999, D. Silvestrov has got a Professor position at the Mälardalen University (Västerås). Here, D. Silvestrov initiated new advanced bachelor and master programs in the area of financial engineering and began an intensive research in the area of stochastic approximation methods for modulated price processes and American-type options. The recent two volumes comprehensive monograph [12, 13] represents the main his results in this area.

In 2009, D. Silvestrov got a prestigious Cramér Professor position at the Department of Mathematics, Stockholm University. In 2013, the International Cramér Symposium on Insurance Mathematics was initiated by D. Silvestrov and held at Stockholm University. The collective book [11] includes a representative sample of papers presented at this symposium. The book [14], co-authored with Professor S. Silvestrov, represents the current research interests of D. Silvestrov in the area of asymptotic expansions for nonlinearly and singularly perturbed semi-Markov type processes and their applications to stochastic networks.

A pedagogical work is an important part of academic work. During his long carrier D. Silvestrov delivered more than 40 different courses on the theory of probability, stochastic processes, statistical software, financial and insurance mathematics, etc. He supervised more than 60 diploma works; 22 postgraduate students (M. Petersson, E. Ekheden (co-supervised with Professor O. Hössjer), Y. Ni, R. Lundgren, M. Drozdenko, F. Stenberg, H. Jönsson, E. Englund, Ö. Stenflo (co-supervised with Professor H. Wallin), Z. Abadov, G. Pezhinska-Pozdnjakova, D. Korolyuk, Yu. Khusambaev, A. Motsa, D. Banakh, E. Kaplan, Yu. Mishura, N. Kartashov, V. Poleshchuk, G. Tursunov, R. Mileshina, and V. Masol) were all supervised by D. Silvestrov and obtained Ph.D. equivalent degrees. Three of them, N. Kartashov, V. Masol, and Yu. Mishura, later became Professors.

During the 50 years of intensive research works D. Silvestrov published 11 books, more than 150 research papers, and co-edited 15 collective works in the area of stochastic processes and their applications.

A more detailed account of Professor D. Silvestrov's academic activities can be found at his web-page: http://www.su.se/profiles/dsilv/.

References

1. Silvestrov, D.S.: Limit Theorems for Semi-Markov Processes and their Applications to Random Walks. Candidate of Science dissertation. Kiev State University, p. 205 (1969)
2. Silvestrov, D.S.: Limit Theorems for Composite Random Functions. Doctor of Science dissertation. Kiev State University, p. 395 (1972)
3. Silvestrov, D.S.: Limit Theorems for Composite Random Functions, p. 318. Vishcha Shkola and Izdatel'stvo Kievskogo Universiteta (1974)
4. Dorogovtsev, A.Ya., Silvestrov, D.S., Skorokhod, A.V., Yadrenko, M.I.: Probability Theory: Collection of Problems, p. 384. Vishcha Shkola, Kiev (1976) (2nd extended edition: Vishcha Shkola, Kiev, 1980, p. 432. English translation: Probability Theory: Collection of Problems. Translations of mathematical monographs, vol. 163, p. xii+347. American Mathematical Society, Providence, RI (1997))
5. Silvestrov, D.S.: Semi-Markov Processes with a Discrete State Space. Library for an Engineer in Reliability, p. 272. Sovetskoe Radio, Moscow (1980)
6. Silvestrov, D.S.: A Software of Applied Statistics, p. 240. Finansi and Statistika, Moscow (1988)
7. Sillvestrov, D.S., Semenov, N.A., Marishchuk, V.V.: Packages of Applied Programs of Statistical Analysis, p. 174. Tekhnika, Kiev (1990)
8. Silvestrov, D.S., Silvestrova, E.D.: Elsevier's Dictionary of Statistical Terminology. English-Russian, Russian-English, p. 496. Elsevier, Amsterdam (1995)
9. Silvestrov D.S.: Limit Theorems for Randomly Stopped Stochastic Processes. Probability and its applications, p. xvi+398. Springer, London (2004)
10. Gyllenberg, M., Silvestrov, D.S.: Quasi-Stationary Phenomena in Nonlinearly Perturbed Stochastic Systems. De Gruyter expositions in mathematics, vol. 44, p. ix+579. Walter de Gruyter, Berlin (2008)
11. Silvestrov, D., Martin-Löf, A. (eds.): Modern Problems in Insurance Mathematics. European actuarial academy (EAA) series, p. xvii+385. Springer, Cham (2014)
12. Silvestrov D.S.: American-Type Options. Stochastic Approximation Methods, Volume 1. De Gruyter studies in mathematics, vol. 56, p. x+509. Walter de Gruyter, Berlin (2014)
13. Silvestrov, D.S.: American-Type Options. Stochastic Approximation Methods, Volume 2. De Gruyter studies in mathematics, vol. 57, p. xi+558. Walter de Gruyter, Berlin (2015)
14. Silvestrov, D., Silvestrov, S.: Nonlinearly Perturbed Semi-Markov Processes. Springer briefs in probability and mathematical statistics, p. xiv+143. Springer, Cham (2017)

Part I
Stochastic Processes

Chapter 2
A Journey in the World of Stochastic Processes

Dmitrii Silvestrov

Abstract This paper presents a survey of research results obtained by the author and his collaborators in the areas of limit theorems for Markov-type processes and randomly stopped stochastic processes, renewal theory and ergodic theorems for perturbed stochastic processes, quasi-stationary distributions for perturbed stochastic systems, methods of stochastic approximation for price processes, asymptotic expansions for nonlinearly perturbed semi-Markov processes and applications of the above results to queuing systems, reliability models, stochastic networks, bio-stochastic systems, perturbed risk processes, and American-type options.

Keywords Limit theorem · Markov-type process · Random stopping · Perturbed renewal equation · Coupling · Quasi-stationary distribution · American type option · Perturbed semi-Markov process

2.1 Introduction

This paper presents a survey of research results in the area of stochastic processes obtained by me and my collaborators during a long period, which began about 50 years ago. My first results have been published in papers [1, 2]. The corresponding complete bibliography of works on stochastic processes and related areas includes 11 books, more than 150 research papers and 15 editorial works. It can be found at the webpage [68].

The main areas of research cover limit theorems for Markov-type processes and randomly stopped stochastic processes, renewal theory and ergodic theorems for perturbed stochastic processes, quasi-stationary distributions for perturbed stochastic systems, methods of stochastic approximation for price processes, asymptotic expansions for nonlinearly perturbed semi-Markov processes, and their applications to queuing systems, reliability models, stochastic networks, bio-stochastic systems,

D. Silvestrov (✉)
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
e-mail: silvestrov@math.su.se

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_2

perturbed risk processes, and American-type options. For convenience of readers, the works pointed in the references are ordered by years of their publication. This makes it possible to see all together works related to every research direction mentioned above.

I would also like to mention paper [61] and books [44, 49, 57, 58, 65], which contain comprehensive bibliographies of works in the above research areas, and the corresponding bibliographical remarks with historical and methodological comments.

About half of my works are written or co-edited together with more than 50 collaborators, including more than 20 of my former doctoral students. Their names can be found in the references given in this paper and the complete bibliography given at [68].

I would like to use this opportunity and to sincerely thank all my collaborators for the fruitful cooperation.

This survey was presented at the International Conference "Stochastic Processes and Algebraic Structures – From Theory Towards Applications" (SPAS 2017, https://spas2017blog.wordpress.com), which was organised on the occasion of my 70th birthday and held at Västerås – Stockholm, on 4–6 October 2017.

I am very grateful to the Organising and Scientific Committees, keynote speakers and other conference participants as well as to the Division of Applied Mathematics (School of Education, Culture and Communication, Mälardalen University) and the Department of Mathematics (Stockholm University) who have supported this conference.

2.2 Limit Theorems for Markov-Type Processes

The main objects of research studies in this area were limit theorems for sums and stepwise sum-processes of random variables defined on asymptotically ergodic and asymptotically recurrent random walks, Markov chains, and semi-Markov processes.

It is worth noting that limit theorems for sums of random variables defined on Markov chains are a very natural generalisation of classical limit theorems for sums of independent random variables. In the case of asymptotically ergodic Markov chains, the corresponding conditions of convergence are similar to well-known classical conditions of convergence for sums and sum-processes of i.i.d. random variables. Also, as in the above classical case, Lévy processes appear as limiting ones.

In the case of asymptotically recurrent Markov chains, the corresponding conditions of convergence and limiting processes take much more complex forms. Possible limiting processes have been described for the above sum-processes defined on Markov chains and semi-Markov processes, first, with countable and, then with general phase spaces. The corresponding limiting processes are generalised exceeding processes. Such processes are constructed with the use of two-dimensional, càdlàg Lévy processes with the second nonnegative component, in the following way. The first component of the above Lévy process is randomly stopped at the moment of

first exceeding a level t (which plays the role of time) by the second component. In addition, this stopping process can possibly be truncated by some exponentially distributed random variable (independent of the above two-dimensional Lévy process) taking into account asymptotic recurrence effects for the above mentioned Markov-type processes.

The corresponding asymptotic results have been obtained firstly in the form of usual limit theorems about weak convergence of finite-dimensional distributions for the corresponding sum-processes, and then in the more advanced form of functional limit theorems about convergence of these processes in the uniform U and Skorokhod J topologies.

Further, the above sum-processes, randomly stopped at different Markov moments, such as hitting times for the above asymptotically recurrent Markov-type processes, have been thoroughly studied and analogous asymptotic results have been obtained as well.

The results related to finite and countable Markov chains and semi-Markov-type processes are well presented in paper [4], dissertation [3], based on 14 research papers, and two books, [8, 14].

Later works in this area have been concentrated on limit theorems for Markov and semi-Markov-type models with general phase spaces, [13, 20, 22, 28], finding not only sufficient but also necessary conditions of convergence, [21, 24, 31], as well as generalisation of the above limit theorems to non-Markov models, [11, 17–19, 26].

The latest results in this area concern necessary and sufficient conditions of convergence for first-rare-event times and processes, [47, 60]. It is appropriate to note that these results yield, in particular, necessary and sufficient conditions for diffusion and stable approximations of ruin probabilities for classical risk processes.

2.3 Limit Theorems for Randomly Stopped Stochastic Processes

A natural research area connected with limit theorems for Markov-type processes is that of limit theorems for randomly stopped stochastic processes and for compositions of stochastic processes.

This model can appear in a number of natural ways, for example: when studying limit theorems for additive or extremal functionals of stochastic processes; in models connected with a random change of time, change point problems and problems related to optimal stopping of stochastic processes; and in different renewal models, particularly those which appear in applications to risk processes, queuing systems, etc.

This model also appears in statistical applications connected with studies of samples with a random sample size. Such sample models play an important role in sequential analysis. They also appear in sample survey models, or in statistical models, where sample variables are associated with stochastic flows. The latter

models are typical for insurance, queuing and reliability applications, as well as many others.

There exists a huge bibliography of works devoted to limit theorems for models with independent or asymptotically independent external processes and stopping moments.

The aim of the author was to build a general theory of limit theorems for compositions of dependent càdlàg external and internal non-negative and non-decreasing stopping processes.

Pre-limiting joint distributions of external and stopping processes usually have a complicated structure. The idea was to find conditions of convergence, where these processes would be involved together only in the simplest and most natural way, via the condition of their joint weak convergence. Also, conditions of compactness in Skorokhod J-topology should be required for external processes and internal stopping processes, in order to provide compactness in Skorokhod J-topology for their compositions. These conditions are standard ones. They were thoroughly studied for various classes of càdlàg stochastic processes.

However, it turns out that the above three conditions do not provide convergence in J-topology for compositions. Some additional assumptions, which link discontinuity moments and values of the corresponding limiting processes at these moments, should be made. First, the probability that the limiting internal stopping process takes the same value at two moments of time $t' < t''$ and this value hits the set of discontinuity moments for the limiting external process should be 0, for any $0 \leq t' < t'' < \infty$. Second, the probability that the intersection of the set of left and right limiting values for the limiting internal stopping process at its jump moments and the set of discontinuity moments for the limiting external process is non-empty should be 0.

The limiting processes usually have simpler structure than the corresponding pre-limiting processes. This permits one to check the above continuity conditions in various practically important cases. For example, the first continuity condition holds, if the limiting external process is a.s. continuous or the limiting internal stopping process is a.s. strongly monotonic. The above second continuity condition holds, if at least one of the limiting external or internal processes is a.s. continuous. Also, both conditions hold if the limiting external and internal processes are independent and the limiting external process is stochastically continuous. The above continuity conditions can not be omitted. If the first continuity condition does not hold for some $t' < t''$, the compositions may not weakly converge at interval $[t', t'']$. If the second continuity condition does not hold, the compositions may not be compact in J-topology.

One of the main theorems proven by the author states that the five conditions listed above do imply J-convergence for compositions of càdlàg processes. These conditions have a good balance that makes this theorem a flexible and effective tool for obtaining limit theorems for randomly stopped stochastic processes.

The main new results found by the author include general limit theorems about weak convergence of randomly stopped stochastic processes and compositions of dependent càdlàg stochastic processes, functional limit theorems about convergence of compositions of càdlàg stochastic processes in topologies U and J as well as

applications of these theorems to random sums, extremes with random sample size, generalised exceeding processes, sum-processes with renewal stopping, accumulation processes, max-processes with renewal stopping, and shock processes.

Some of the most valuable works in this area are the following papers: [5, 6, 9, 36, 41–43], dissertation [7], based on 27 research papers, and book [8]. The final extended version of the theory developed by the author is presented in the book [44].

2.4 Renewal Theory and Ergodic Theorems for Perturbed Stochastic Processes

Another research area connected in a natural way with limit theorems for Markov-type processes is renewal theory and ergodic theorems for perturbed stochastic processes.

An important role is played in both limit and ergodic theorems by such random functionals as hitting times and their moments. Necessary and sufficient conditions of existence and the most general explicit recurrent upper bounds for power and exponential moments of hitting-time type functionals for semi-Markov processes have been given in book [14] and papers [34, 46].

Further, related recurrent computational algorithms based on various truncation and phase space reduction procedures are given for semi-Markov-type processes and networks in papers [16, 56, 64].

Ergodic theorems of the law of larger numbers type and related ergodic theorems for mean averages for accumulation processes and iterated functions systems are given in papers [15, 31, 37] and book [14].

Uniform asymptotic expansions for exponential moments of sums of random variables defined on exponentially ergodic Markov chains and distributions of hitting times for such Markov chains are given in paper [29].

As is well known, the most effective tool for getting so-called individual ergodic theorems for regenerative and Markov-type processes is the famous renewal theorem. In the case of perturbed processes, an effective generalisation of this important theorem to the model of perturbed renewal equation is required. Such a generalisation was given in paper [12].

These results and their applications to perturbed regenerative, semi-Markov and risk processes, and ergodic theorems for perturbed queuing systems and biostochastic systems are presented in papers [40, 55] and book [49]. Also, the recent paper, [66], presents results of a detailed analysis and classification of ergodic theorems for perturbed alternating regenerative processes.

In papers [23, 30], exact coupling algorithms have been composed for general regenerative processes and stochastic processes with semi-Markov modulation, and explicit estimates for the rate of convergence in related individual ergodic theorems for such processes were given. It is worth noting that, in the continuous time case, the algorithms of exact coupling do require construction of dependent coupling

trajectories in the essentially more sophisticated way, if we are to compare them with the corresponding coupling algorithms for discrete time processes.

In addition, paper [53] can be mentioned, where the above coupling algorithms are applied for obtaining explicit estimates for the rate of convergence in the classical Cramér-Lundberg approximation for ruin probabilities.

2.5 Quasi-Stationary Phenomena for Perturbed Stochastic Systems

Quasi-stationary phenomena in stochastic systems describe the behaviour of stochastic systems with random lifetimes. The core of the quasi-stationary phenomenon is that one can observe something that resembles a stationary behaviour of the system before the lifetime goes to the end.

The objects of interest are the asymptotic behaviour of lifetimes in the forms of weak convergence and large deviation theorems, conditional ergodic theorems (describing the asymptotic behaviour, when $t \to \infty$, for the conditional distribution of the corresponding stochastic process at moment t, under the condition that the lifetime takes a value larger than t), and the corresponding limiting, usually referred to as quasi-stationary, distributions.

In the model of perturbed stochastic processes of Markov-type, their transition characteristics depend on a small perturbation parameter ε and, moreover, they may admit some asymptotic expansions with respect to this parameter. A problem arises in constructing asymptotic expansions for distributions of lifetimes, conditional distributions for the underlying stochastic processes pointed out above, and the corresponding quasi-stationary distributions.

It is relevant to note that quasi-stationary distributions are essentially nonlinear functionals of transition characteristics for underlying Markov-type processes. They depend on so-called characteristic roots for distributions of return times for the above processes. This significantly complicates the problem of constriction of asymptotic expansions for quasi-stationary distributions, if we are to compare it with the analogous problem for ordinary stationary distributions of perturbed Markov-type processes.

It also turns out that the balance between the velocities with which ε tends to zero and time t tends to infinity (expressed in the form of asymptotic relation, $t\varepsilon^r \to \lambda_r \in [0, \infty]$) has a delicate influence on the quasi-stationary asymptotics. The above-mentioned expansions make it possible to perform the corresponding detailed asymptotic analysis.

New methods based on asymptotic expansions for solutions of perturbed renewal equations have been proposed in paper [32] and used in papers [38, 39, 55] for finding quasi-stationary asymptotics given in the form of asymptotic expansions, for perturbed regenerative processes, Markov chains, semi-Markov processes, and risk processes. Also, asymptotic expansions for discrete time nonlinearly perturbed

renewal equations have been given in papers [35, 54] and for the renewal equation with nonlinear non-polynomial perturbations in paper [48].

The comprehensive book [49] contains a detailed presentation of the above mentioned methods for nonlinearly perturbed regenerative processes and finite semi-Markov processes with absorption. It also includes their applications to the analysis of quasi-stationary phenomena in nonlinearly perturbed highly reliable queuing systems, M/G queuing systems with quick service, and stochastic systems of birth-death type, including perturbed epidemic, population and meta-population models, and perturbed risk processes.

Also, the paper [66] presents new ergodic theorems for perturbed alternating regenerative processes obtained with the use of quasi-stationary ergodic theorems for perturbed regenerative processes.

2.6 Stochastic Approximation Methods for Price Processes and American-Type Options

American-type options are one of the most important financial instruments and, at the same time, one of the most interesting and popular objects for research studies in financial mathematics.

The main mathematical problems connected with such options relate to finding of the optimal expected option rewards, in particular, fair prices of options, as well as finding of optimal strategies for buyers of options that are optimal stopping times for execution of options.

In this way, the theory of American-type options is connected with optimal stopping problems for stochastic processes, which play an important role in the theory of stochastic processes and its applications.

As is well known, analytical solutions for American-type options are available only in some special cases and, even in such cases, the corresponding formulas are not easily computable. These difficulties dramatically increase in the case of multivariate log-price processes and non-standard pay-off functions.

Approximation methods are a reasonable alternative that can be used in cases where analytical solutions are not available. The main classes of approximation methods are: stochastic approximation methods based on approximation of the corresponding stochastic log-price processes by simpler processes, for which optimal expected rewards can be effectively computed; integro-differential approximation methods based on approximation of integro-differential equations that can be derived for optimal expected rewards by their difference analogues; and Monte Carlo methods based on simulation of the corresponding log-price processes.

Stochastic approximation methods have important advantages in comparison with the other two methods. They usually allow one to impose weaker smoothness conditions on transition probabilities and pay-off functions, in comparison with

integro-differential approximation methods, and they are also computationally more effective, in comparison with Monte Carlo based methods.

Selected papers presenting the author and his collaborators' results in the above area are [45, 50–52, 59]. Some of these results, as well as many new results, are presented in the author's comprehensive two volume monograph, [57, 58].

This monograph gives a systematic presentation of stochastic approximation methods for models of American-type options with general pay-off functions for discrete (Volume 1) and continuous (Volume 2) time modulated Markov log-price processes. Advanced methods, combining backward recurrence algorithms for computing of option rewards for discrete time atomic Markov chains with transition probabilities concentrated on finite sets, and general results on convergence of the corresponding stochastic time-space skeleton and tree approximations for option rewards, are applied to a variety of models of multivariate modulated Markov log-price processes.

In the discrete time case, these are modulated autoregressive and autoregressive stochastic volatility log-price processes, log-price processes represented by modulated random walks, Markov Gaussian log-price processes with estimated parameters, multivariate modulated Markov Gaussian log-price processes and their binomial and trinomial approximations, and log-price processes represented by general modulated Markov chains.

In the continuous time case, these are multivariate Lévy log-price processes, multivariate diffusion log-price processes and their time-skeleton, martingale and trinomial approximations, and general continuous time multivariate modulated Markov log-price processes.

Further, some more advanced models of American-type options are treated; in particular, options with random pay-offs, reselling options and knockout options.

The principal novelty of results presented in the monograph [57, 58] is based on the consideration of multivariate modulated Markov log-price processes and general pay-off functions, which can depend not only on price but also on an additional stochastic modulating index component, and the use of minimal conditions of smoothness for transition probabilities and pay-off functions, compactness conditions for log-price processes and rate of growth conditions for pay-off functions.

2.7 Nonlinearly Perturbed Semi-Markov Processes

The models of perturbed Markov chains and semi-Markov processes attracted the attention of researchers in the middle of the twentieth century. Particular attention was given to the most difficult cases of perturbed processes with absorption and the so-called singularly perturbed processes. An interest in these models has been stimulated by applications to control and queuing systems, reliability models, information networks, and bio-stochastic systems.

Markov-type processes with singular perturbations appear as natural tools for mathematical analysis of multicomponent systems with weakly interacting

components. Asymptotics for moments of hitting-time type functionals and stationary distributions for corresponding perturbed processes play an important role in studies of such systems.

The role of perturbation parameters can be played by small probabilities or failure rates in queuing and reliability systems, or by small probabilities or intensities of mutation, extinction, or migration in biological systems. Perturbation parameters can also appear as artificial regularisation parameters for decomposed systems, for example, as so-called damping parameters in information networks, etc.

In many cases, transition characteristics of the corresponding perturbed semi-Markov processes, in particular transition probabilities (of embedded Markov chains) and power moments of transition times are nonlinear functions of a perturbation parameter, which admit asymptotic expansions with respect to this parameter.

The main results obtained so far in these ongoing studies are presented in papers [61–63] and the recent book [65].

These works present new methods of asymptotic analysis for nonlinearly perturbed semi-Markov processes with finite phase spaces. These methods are based on special time-space screening procedures for sequential reduction of phase spaces for semi-Markov processes combined with the systematic use of the operational calculus for Laurent asymptotic expansions.

Models with non-singular and singular perturbations are considered to be those where the phase space is one class of communicative states for the embedded Markov chains of pre-limiting perturbed semi-Markov processes, while it can possess an arbitrary communicative structure (i.e., can consist of one or several closed classes of communicative states and, possibly, a class of transient states) for the limiting embedded Markov chain.

Effective recurrent algorithms for the construction of Laurent asymptotic expansions for power moments of hitting times for nonlinearly perturbed semi-Markov processes are composed. These results are applied for obtaining asymptotic expansions for stationary and conditional quasi-stationary distributions for nonlinearly perturbed semi-Markov processes. Also, the detailed asymptotic analysis and the corresponding asymptotic expansions are given for semi-Markov birth-death-type processes, which play an important role in various applications.

Further, the recent paper [67], which presents applications of asymptotic expansions for semi-Markov birth-death-type processes to perturbed models of population dynamics, epidemic models and models of population genetics, should be mentioned.

It is worth noting that asymptotic expansions are a very effective instrument for studies of perturbed stochastic processes. The corresponding first terms in expansions give limiting values for properly normalised functionals of interest. The second terms let one estimate the sensitivity of models to small parameter perturbations. The subsequent terms in the corresponding expansions are usually neglected in standard linearisation procedures used in studies of perturbed models. This, however, cannot be acceptable in cases where values of perturbation parameters are not small enough. Asymptotic expansions let one take into account high-order terms in expansions, and in this way allow one to improve accuracy of the corresponding numerical procedures.

An important novelty of the results presented in works [61–63, 65] is that the corresponding asymptotic expansions are obtained with remainders given not only in the standard form of $o(\varepsilon^k)$, but, also, in the more advanced form, with explicit power-type upper bounds for remainders, $|o(\varepsilon^k)| \le G_k \varepsilon^{k+\delta_k}$, asymptotically uniform with respect to the perturbation parameter. The latter asymptotic expansions for nonlinearly perturbed semi-Markov processes were not known before.

The corresponding computational algorithms have a universal character. They can be applied to perturbed semi-Markov processes with an arbitrary asymptotic communicative structure of phase spaces and are computationally effective due to the recurrent character of computational procedures.

2.8 Conclusion

There exist a number of prospective directions for continuation of research and many interesting unsolved problems in the research areas listed above.

In the area of limit theorems for Markov-type processes, the methods developed so far, particularly general limit theorems for randomly stopped stochastic processes, allow, I believe, to obtain new, more advanced versions of these limit theorems and, moreover, to improve conditions of convergence for some of these theorems to the final necessary and sufficient form, without gaps between necessary and sufficient counterparts. One of the most recently published papers, [60], provides an example of such results.

The theory of limit theorems for randomly stopped stochastic processes can also be effectively used in such applied areas as asymptotic problems of statistical analysis. The book [44] contains some results related to samples with random size. However, this very prospective area of applications is still underdeveloped.

In the area of renewal theory and ergodic theorems for perturbed Markov-type processes, applications of the exact coupling method presented in papers [23, 30] can be essentially extended. For example, the results of the above-mentioned paper [53] show how this method can be effectively applied to risk processes.

The book [49] contains a survey of new potential directions of studies related to quasi-stationary phenomena for perturbed stochastic systems. I believe it will be possible to combine asymptotic quasi-stationary and coupling methods for perturbed renewal equations and to obtain explicit upper bounds for rates of convergence in the corresponding limit, large deviation and ergodic theorems. In this book, results on quasi-stationary asymptotics for perturbed regenerative processes are applied to models of nonlinearly perturbed Markov chains and semi-Markov processes with finite phase spaces. An analogous program of research studies can be realised for Markov chains and semi-Markov type processes with countable and general phase spaces, which also can be embedded in the class of regenerative processes using the well known method of artificial regeneration. Another direction for advancing the studies carried out in this book is connected with models of nonlinearly perturbed

Markov chains and semi-Markov processes with asymptotically uncoupled phase spaces. One of the latest my paper [66] presents new results in this research direction.

Stochastic approximation methods are, as was pointed out above, one of the most effective instruments for studies of complex financial contracts. Time-space skeleton approximation methods developed in the books [57, 58] can be applied to American-type contracts of different types. In particular, the above-mentioned results for American-type options with estimated parameters, options with random pay-offs, reselling options and knockout options given in the above books can be essentially extended. Also, these methods can be effectively applied to European and Asian options, as well as Bermudian and other types of exotic options. Another prospective area is associated with the combination of the above stochastic approximation methods with Monte Carlo type algorithms.

The latest actual direction of research of the author and some of his collaborators are asymptotic expansions for nonlinearly and singularly perturbed semi-Markov processes. The method of sequential phase space reduction proposed in the works [61–63, 65] allows, as was mentioned above, for the attainment of asymptotic expansions for different moment functionals for non-singularly and singularly perturbed semi-Markov process in two forms, without and with explicit upper bounds for remainders.

Both the class of semi-Markov type processes and the class of moment functionals can be essentially extended. For example, asymptotic expansions for power and power-exponential moments of hitting times and quasi-stationary distributions can be obtained for singularly perturbed stochastic processes with semi-Markov modulation.

Prospective directions of research studies based on the above methods are asymptotic expansions for perturbed semi-Markov type processes with multivariate perturbation parameters and asymptotic expansions based on non-polynomial systems of infinitesimals.

Further, an unbounded area of applications is perturbed queuing and reliability models, stochastic networks and bio-stochastic systems.

In conclusion, I would also like to mention some books which reflect my interests in some other scientific areas thematically connected with stochastic processes and their applications, [10, 25, 27, 33].

My journey in the world of stochastic processes continues.

References

1. Silvestrov, D.S.: A generalisation of Pólya's theorem. Dokl. Akad. Nauk Ukr. SSR, Ser A (10), 906–909 (1968)
2. Silvestrov, D.S.: Limit theorems for a non-recurrent walk connected with a Markov chain. Ukr. Math. Zh. **21**, 790–804 (1969). (English translation in Ukr. Math. J. **21**, 657–669)
3. Silvestrov, D.S.: Limit Theorems for Semi-Markov Processes and their Applications to Random Walks. Candidate of Science dissertation, Kiev State University, 205 pp. (1969)

4. Silvestrov, D.S.: Limit theorems for semi-Markov processes and their applications. 1, 2. Teor. Veroyatn. Math. Stat. **3**; Part 1: 155–172; Part 2: 173–194 (1970) (English translation in Theory Probab. Math. Stat. **3**, Part 1: 159–176; Part 2: 177–198)
5. Silvestrov, D.S.: Remarks on the limit of composite random function. Teor. Veroyatn. Primen. **17**, 707–715 (1972). (English translation in Theory Probab. Appl. **17**, 669–677)
6. Silvestrov, D.S.: The convergence of weakly dependent processes in the uniform topology. 1, 2. Teor. Veroyatn. Math. Stat. Part 1: **6**, 104–117; Part 2: **7**, 132–145 (1972) (English translation in Theory Probab. Math. Stat. Part 1: **6**, 109–119; Part 2: **7**, 125–138)
7. Silvestrov, D.S.: Limit Theorems for Composite Random Functions. Doctor of Science dissertation, Kiev State University, 395 pp. (1972)
8. Silvestrov, D.S.: Limit Theorems for Composite Random Functions, 318 pp. Vysshaya Shkola and Izdatel'stvo Kievskogo Universiteta, Kiev (1974)
9. Silvestrov, D.S., Mirzahmedov, M.A., Tursunov, G.T.: On the applications of limit theorems for composite random functions to certain problems in statistics. Theor. Veroyatn. Math. Stat. **14**, 124–137 (1976) (English translation in Theory Probab. Math. Stat. **14**, 133–147)
10. Dorogovtsev, A.Yu., Silvestrov, D.S. Skorokhod, A.V., Yadrenko, M.I.: Probability Theory. A Collection of Problems. Vishcha Shkola, Kiev, 384 pp. (1976) (2nd extended edition: Vishcha Shkola, Kiev, 432 pp. (1980); English translation: Probability Theory: Collection of Problems. Translations of mathematical monographs, vol. 163, 347 pp. American Mathematical Society (1997))
11. Silvestrov D.S., Tursunov, G.T.: General limit theorems for sums of controlled random variables. 1, 2. Teor. Veroyatn. Math. Stat. Part 1: **17**, 120–134 (1977); Part 2: **20**, 116–126 (1979) (English translation in Theory Probab. Math. Stat. Part 1: **17**, 131–146; Part 2: **20**, 131–141)
12. Silvestrov, D.S.: The renewal theorem in a series scheme. 1, 2. Teor. Veroyatn. Math. Stat. Part 1: **18**, 144–161 (1978); Part 2: **20**, 97–116 (1979) (English translation in Theory Probab. Math. Stat. Part 1: **18**, 155–172; Part 2: **20**, 113–130)
13. Kaplan, E.I., Silvestrov, D.S.: Theorems of the invariance principle type for recurrent semi-Markov processes with arbitrary phase space. Teor. Veroyatn. Primen. **24**, 529–541 (1979) (English translation in Theory Probab. Appl. **24**, 537–547)
14. Silvestrov, D.S.: Semi-Markov Processes with a Discrete State Space. Library for an Engineer in Reliability, 272 pp. Sovetskoe Radio, Moscow (1980)
15. Silvestrov, D.S.: Remarks on the strong law of large numbers for accumulation processes. Teor. Veroyatn. Math. Stat. **22**, 118–130 (1980) (English translation in Theory Probab. Math. Stat. **22**, 131–143)
16. Silvestrov, D.S.: Mean hitting times for semi-Markov processes and queueing networks. Elektronische Inf. Kybern. **16**, 399–415 (1980)
17. Kaplan, E.I., Silvestrov, D.S.: General limit theorems for sums of controlled random variables with an arbitrary state space for the controlling sequence. Litov. Math. Sbornik, **20**(4), 61–72 (1980) (English translation in Lith. Math. J. **20**, 286–293)
18. Silvestrov, D.S.: Theorems of large deviations type for entry times of a sequence with mixing. Teor. Veroyatn. Math. Stat. **24**, 129–135 (1981) (English translation in Theory Probab. Math. Stat. **24**, 145–151)
19. Silvestrov, D.S., Khusanbayev, Ya.M.: General limit theorems for random processes with conditionally independent increments. Theor. Veroyatn. Math. Stat. **27**, 130–139 (1982) (English translation in Theory Probab. Math. Stat. **27**, 147–155)
20. Kaplan, E.I., Motsa, A.I., Silvestrov, D.S.: Limit theorems for additive functionals defined on asymptotically recurrent Markov chains. 1, 2. Teor. Veroyatn. Math. Stat. Part 1: **27**, 34–51 (1982); Part 2: **28**, 31–40 (1983) (English translation in Theory Probab. Math. Stat. Part 1: **27**, 39–54; Part 2: **28**, 35–43)
21. Korolyuk, D.V., Silvestrov D.S.: Entry times into asymptotically receding domains for ergodic Markov chains. Teor. Veroyatn. Primen. **28**, 410–420 (1983) (English translation in Theory Probab. Appl. **28**, 432–442)

22. Silvestrov, D.S.: Invariance principle for the processes with semi-Markov switch-overs with an arbitrary state space. In: Itô, K., Prokhorov, YuV (eds.) Proceedings of the Fourth USSR–Japan Symposium on Probability Theory and Mathematical Statistics, Tbilisi, 1982. Lecture notes in mathematics, vol. 1021, 617–628 pp. Springer, Berlin (1983)
23. Silvestrov, D.S.: Method of a single probability space in ergodic theorems for regenerative processes. 1 – 3. Math. Operat. Stat. Ser. Optim. Part 1: **14**, 285–299 (1983); Part 2: **15**, 601–612 (1984); Part 3: **15**, 613–622 (1984)
24. Silvestrov, D.S., Velikiĭ, YuA: Necessary and sufficient conditions for convergence of attainment times. In: Zolotarev, V.M., Kalashnikov, V.V. (eds.) Stability Problems for Stochastic Models. Trudy Seminara, 129–137 pp. VNIISI, Moscow (1988). (English translation in J. Soviet. Math. **57**, 3317–3324)
25. Silvestrov, D.S.: A Software of Applied Statistics, 240 pp. Finansi and Statistika, Moscow (1988)
26. Sillvestrov, D.S., Brusilovskiĭ, I.L.: An invariance principle for sums of stationary connected random variables satisfying a uniform strong mixing condition with a weight. Teor. Veroyatn. Math. Stat. **40**, 99–107 (1989) (English translation in Theory Probab. Math. Stat. **40**, 117–127)
27. Silvestrov, D.S., Semenov, N.A., Marishchuk, V.V.: Packages of applied programs of statistical analysis, 174 pp. Kiev, Tekhnika (1990)
28. Sillvestrov, D.S.: The invariance principle for accumulation processes with semi-Markov switchings in a scheme of arrays. Teor. Veroyatn. Primen. **36**(3), 505–520 (1991) (English translation in Theory Probab. Appl. **36**(3), 519–535)
29. Silvestrov, D.S., Abadov, Z.A.: Uniform representations of exponential moments of sums of random variables defined on a Markov chain and for distributions of passage times. 1, 2. Teor. Veroyatn. Math. Stat. Part 1: **45**, 108–127 (1991); Part 2: **48**, 175–183 (1993) (English translation in Theory Probab. Math. Stat. Part 1: **45**, 105–120; Part 2: **48**, 125–130)
30. Silvestrov, D.S.: Coupling for Markov renewal processes and the rate of convergence in ergodic theorems for processes with semi-Markov switchings. Acta Appl. Math. **34**, 109–124 (1994)
31. Velikiĭ, Yu.A., Motsa, A.I., Silvestrov, D.S.: Dual processes and ergodic type theorems for Markov chains in the triangular array scheme. Teor. Veroyatn. Primen. **39**, 716–730 (1994) (English translation in Theory Probab. Appl. **39**, 642–653)
32. Silvestrov, D.S.: Exponential asymptotic for perturbed renewal equations. Teor. Ĭmovirn. Math. Stat. **52**, 143–153 (1995) (English translation in Theory Probab. Math. Stat. **52**, 153–162)
33. Silvestrov, D.S., Silvestrova, E.D.: Elsevier's Dictionary of Statistical Terminology. English-Russian, Russian-English, 496 pp. Elsevier, Amsterdam (1995)
34. Silvestrov, D.S.: Recurrence relations for generalised hitting times for semi-Markov processes. Ann. Appl. Probab. **6**, 617–649 (1996)
35. Englund, E., Silvestrov, D.: Mixed large deviation and ergodic theorems for regenerative processes with discrete time. In: Jagers, P., Kulldorff, G., Portenko, N., Silvestrov, D. (eds.) Proceedings of the Second Scandinavian–Ukrainian Conference in Mathematical Statistics, vol. I, Umeå (1997). Theory Stoch. Process. **3(19)**(1–2), 164–176 (1997)
36. Silvestrov, D.S., Teugels, J.L.: Limit theorems for extremes with random sample size. Adv. Appl. Probab. **30**, 777–806 (1998)
37. Silvestrov, D.S., Stenflo, Ö.: Ergodic theorems for iterated function systems controlled by regenerative sequences. J. Theor. Probab. **11**, 589–608 (1998)
38. Gyllenberg, M., Silvestrov, D.S.: Nonlinearly perturbed regenerative processes and pseudo-stationary phenomena for stochastic systems. Stoch. Process. Appl. **86**, 1–27 (2000)
39. Gyllenberg, M., Silvestrov, D.S.: Cramér-Lundberg approximation for nonlinearly perturbed risk processes. Insur. Math. Econ. **26**, 75–90 (2000)
40. Silvestrov, D.S.: A perturbed renewal equation and an approximation of diffusion type for risk processes. Teor. Ĭmovirn. Math. Stat. **62**, 134–144 (2000) (English translation in Theory Probab. Math. Stat. **62**, 145–156)
41. Silvestrov, D.S.: Generalized exceeding times, renewal and risk processes. Theory Stoch. Process. **6(22)**(3–4), 125–182 (2000)

42. Silvestrov, D.S., Teugels, J.L.: Limit theorems for mixed max-sum processes with renewal stopping. Ann. Appl. Probab. **14**(4), 1838–1868 (2004)
43. Mishura, Yu.S., Silvestrov, D.S.: Limit theorems for stochastic Riemann-Stieltjes integrals. Theory Stoch. Process. **10(26)**(1–2), 122–140 (2004)
44. Silvestrov D.S.: Limit Theorems for Randomly Stopped Stochastic Processes. Probability and its applications, xiv+398 pp. Springer, London (2004)
45. Jönsson, H., Kukush, A.G., Silvestrov, D.S.: Threshold structure of optimal stopping strategies for American type option. 1, 2. Theor. Ĭmovirn. Math. Stat. Part 1: **71**, 82–92 (2004); Part 2: **72**, 42–53 (2005) (English translation in Theory Probab. Math. Stat. Part 1: **71**, 93–103; Part 2: **72**, 47–58)
46. Silvestrov, D.S.: Upper bounds for exponential moments of hitting times for semi-Markov processes. Commun. Stat. Theory Methods **33**(3), 533–544 (2005)
47. Silvestrov, D.S., Drozdenko, M.O.: Necessary and sufficient conditions for weak convergence of first-rare-event times for semi-Markov processes. 1, 2. Theory Stoch. Process. **12(28)**(3–4); Part 1: 151–186; Part 2: 187–202 (2006)
48. Ni, Y., Silvestrov, D., Malyarenko, A.: Exponential asymptotics for nonlinearly perturbed renewal equation with non-polynomial perturbations. J. Numer. Appl. Math. **1**(96), 173–197 (2008)
49. Gyllenberg, M., Silvestrov, D.S.: Quasi-Stationary Phenomena in Nonlinearly Perturbed Stochastic Systems. De Gruyter expositions in mathematics, vol. 44, ix+579 pp. Walter de Gruyter, Berlin (2008)
50. Silvestrov, D., Jönsson, H., Stenberg, F.: Convergence of option rewards for Markov type price processes modulated by stochastic indices. 1, 2. Theor. Ĭmovirn. Math. Stat. Part 1: **79**, 149–165 (2008); Part 2: **80**, 138–155 (2009) (Also in Theory Probab. Math. Stat. Part 1: **79**, 153–170; Part 2: **80**, 153–172)
51. Lundgren, R., Silvestrov, D.S.: Optimal stopping and reselling of European options. In: Rykov, V., Balakrishan, N., Nikulin, M. (eds.) Mathematical and Statistical Models and Methods in Reliability, Chap. 29, 371–390. Birkhäuser, New York (2010)
52. Silvetsrov, D., Lundgren, R.: Convergence of option rewards for multivariate price processes. Theor. Ĭmovirn. Math. Stat. **85**, 102–116 (2011) (Also in Theory Probab. Math. Stat. **85**, 115–131)
53. Ekheden, E., Silvestrov, D.: Coupling and explicit rates of convergence in Cramér-Lundberg approximation for reinsurance risk processes. Commun. Stat. Theory Methods **40**(19–20), 3524–3539 (2011)
54. Silvestrov, D., Petersson, M.: Exponential expansions for perturbed discrete time renewal equations. In: Karagrigoriou, A., Lisnianski, A., Kleyner, A., Frenkel, I. (eds.) Applied Reliability Engineering and Risk Analysis. Probabilistic Models and Statistical Inference, Chap. 23, 349–362. Wiley, Chichester (2013)
55. Silvestrov, D.: Improved asymptotics for ruin probabilities. In: Silvestrov, D., Martin-Löf, A. (eds.) Modern Problems in Insurance Mathematics, Chap. 5, 37–68. EAA series, Springer, Cham (2014)
56. Silvestrov, D., Manca, R., Silvestrova, E.: Computational algorithms for moments of accumulated Markov and semi-Markov rewards. Commun. Stat. Theory Methods **43**(7), 1453–1469 (2014)
57. Silvestrov D.S.: American-Type Options. Stochastic Approximation Methods, Volume 1. De Gruyter studies in mathematics, vol. 56, x+509 pp. Walter de Gruyter, Berlin (2014)
58. Silvestrov D.S.: American-Type Options. Stochastic Approximation Methods, Volume 2. De Gruyter studies in mathematics, vol. 57, xi+558 pp. Walter de Gruyter, Berlin (2015)
59. Silvestrov, D., Li, Y.: Stochastic approximation methods for American type options. Commun. Stat. Theory Methods **45**(6), 1607–1631 (2016)
60. Silvestrov, D.: Necessary and sufficient conditions for convergence of first-rare-event times for perturbed semi-Markov processes. Theor. Ĭmovirn. Math. Stat. **95**, 119–137 (2016) (Also in Theory Probab. Math. Stat. **95**, 135–151)

61. Silvestrov, D., Silvestrov, S.: Asymptotic expansions for stationary distributions of perturbed semi-Markov processes. In: Silvestrov, S., Rančić, M. (eds.) Engineering Mathematics II. Algebraic, Stochastic and Analysis Structures for Networks, Data Classification and Optimization. Springer proceedings in mathematics and statistics, vol. 179, Chap. 10, 151–222, Springer. Cham (2016)
62. Silvestrov, D., Silvestrov, S.: Asymptotic expansions for stationary distributions of nonlinearly perturbed semi-Markov processes. 1, 2. Method. Comput. Appl. Probab. Part 1: https://doi.org/10.1007/s11009-017-9605-0, 20 pp.; Part 2: https://doi.org/10.1007/s11009-017-9607-y, 20 pp. (2017)
63. Silvestrov, D., Silvestrov, S.: Asymptotic expansions for power-exponential moments of hitting times for nonlinearly perturbed semi-Markov processes. Theor. Ĭmovirn. Math. Stat. **97**, 171–187 (2017)
64. Silvestrov, D., Manca, R.: Reward algorithms for semi-Markov processes. Method. Comput. Appl. Probab. **19**(4), 1191–1209 (2017)
65. Silvestrov, D., Silvestrov, S.: Nonlinearly Perturbed Semi-Markov Processes. Springer briefs in probability and mathematical statistics, xiv+143 pp. Springer, Cham (2017)
66. Silvestrov, D.: Individual ergodic theorems for perturbed alternating regenerative processes. In: Silvestrov, S., Rančić, M., Malyarenko, A. (eds.) Stochastic Processes and Applications. Springer proceedings in mathematics & statistics, vol. 271, Chap. 3. Springer, Cham (2018)
67. Silvestrov, D., Petersson, M., Hössjer, O.: Nonlinearly perturbed birth-death-type models. In: Silvestrov, S., Rančić, M., Malyarenko, A. (eds.) Stochastic Processes and Applications. Springer proceedings in mathematics & statistics, vol. 271, Chap. 11. Springer, Cham (2018)
68. http://www.su.se/profiles/dsilv/

Chapter 3
Individual Ergodic Theorems for Perturbed Alternating Regenerative Processes

Dmitrii Silvestrov

Abstract The paper presents results of complete analysis and classification of individual ergodic theorems for perturbed alternating regenerative processes with semi-Markov modulation. New short, long and super-long time ergodic theorems for regularly and singular type perturbed alternating regenerative processes are presented.

Keywords Alternating regenerative process · Semi-Markov modulation · Regular perturbation · Singular perturbation · Ergodic theorem

3.1 Introduction

The paper presents results of complete analysis and classification of individual ergodic theorems for perturbed alternating regenerative processes with semi-Markov modulation.

The alternating regenerative processes and related alternating renewal processes are popular models of stochastic processes, which have diverse applications to queuing, reliability, control and many other types of stochastic processes and systems. We refer here to papers and books, which contain basic materials about regenerative processes including their alternating variants and applications, [4, 7, 9, 15, 17, 23, 28–32, 46, 55, 59].

Standard alternating regenerative processes are constructed from sequences of "random blocks" of two types, say, 1 and 2. Each block consists of a "piece" of stochastic process of random duration. All blocks are independent. Blocks of each type have the same probabilistic characteristics. The corresponding alternating regenerative process is constructed by sequential in time alternate connection of blocks of types 1 and 2 taken from the above mentioned sequences.

D. Silvestrov (✉)
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
e-mail: silvestrov@math.su.se

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_3

In the present paper, more general alternating regenerative processes are studied, where sequential alternate connection of the blocks is controlled by some binary switching random variables. The piece of stochastic process creating every block, its duration and the binary random variable, controlling the decision about switching/non-switching of block type at the end of time interval corresponding to this block, may be dependent. This let us speak about semi-Markov modulation for the corresponding alternating regenerative process.

If the above alternating regeneration process $\xi_\varepsilon(t), t \geq 0$ describes functioning of some stochastic system, it is naturally to interpret $\xi_\varepsilon(t)$ as the state of this system at instant t and the corresponding modulating semi-Markov process $\eta_\varepsilon(t)$ as the stochastic index, which shows that the system is in one of two possible regimes (for example, "working" or "not working") at instant t if, respectively, $\eta_\varepsilon(t) = 1$ or $\eta_\varepsilon(t) = 2$.

It is assumed that joint probabilistic characteristics of the alternating regenerative process $\xi_\varepsilon(t)$ and the corresponding semi-Markov process $\eta_\varepsilon(t)$ controlling switching of types depend on some perturbation parameter $\varepsilon \in [0, 1]$ and converge to the corresponding joint characteristics of the processes $\xi_0(t)$ and $\eta_0(t)$, as $\varepsilon \to 0$. This makes it possible to consider process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, for $\varepsilon \in (0, 1]$, as a perturbed version of the process $(\xi_0(t), \eta_0(t))$.

The object of our interest are individual ergodic theorems about asymptotic behaviour of joint distributions $\mathsf{P}_{\varepsilon,ij}(t, A) = \mathsf{P}_i\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = j\}$ for perturbed alternating regenerative process $\xi_\varepsilon(t)$ and modulating semi-Markov processes $\eta_\varepsilon(t)$, as time $t \to \infty$ and the perturbation parameter $\varepsilon \to 0$.

Models with three different types of perturbation are considered. These types are determined by the asymptotic behaviour of transition probabilities $p_{\varepsilon,ij}, i, j = 1, 2$ for the embedded Markov chain $\eta_{\varepsilon,n}$ of the semi-Markov process $\eta_\varepsilon(t)$. These transition probabilities converge, as $\varepsilon \to 0$, to the corresponding transition probabilities of the limiting Markov chain $\eta_{0,n}$.

The first class constitutes regularly perturbed models, where the limiting embedded Markov chain $\eta_{0,n}$ is ergodic that, in this case, is equivalent to the assumption,

$$\max(p_{0,12}, p_{0,21}) > 0. \tag{3.1}$$

In the case of regularly perturbed models, the corresponding individual ergodic theorems take forms of asymptotic relations $\mathsf{P}_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_j(A)$ as $\varepsilon \to 0$, which holds for any $0 \leq t_\varepsilon \to \infty$ as $\varepsilon \to 0$. The corresponding limiting probabilities $\pi_j(A)$ do not depend on an initial state i of the modulating semi-Markov process.

Such theorems resemble well known ergodic theorems for unperturbed alternating regenerative processes and more general stochastic processes with semi-Markov modulation.

Here, works [4, 7, 9, 12, 17, 23, 24, 29, 30, 35–37, 46–50, 55, 59] can be referred, where one can find the corresponding ergodic theorems for unperturbed regenerative and alternating regenerative processes $(\xi_0(t), \eta_0(t))$, and works [1, 10, 11, 13, 14, 23, 25–27, 33, 34, 37, 43–45, 47–51, 53, 54, 60–62], where such theorems are given for some classes of regularly perturbed regenerative and alternating regenerative processes $(\xi_\varepsilon(t), \eta_\varepsilon(t))$.

The second and third classes constitute singularly and super-singularly perturbed models, where the limiting embedded Markov chain $\eta_{0,n}$ is not ergodic that is equivalent to the assumption,

$$\max(p_{0,12},\, p_{0,21}) = 0. \tag{3.2}$$

The individual ergodic theorems for such models are the main objects of studies in the present paper. They take much more interesting and complex forms, if to compare them with individual ergodic theorems for regularly perturbed alternating regenerative processes. The individual ergodic theorems for singularly and super-singularly perturbed models take forms of asymptotic relations $\mathsf{P}_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{ij}(t, A)$ as $\varepsilon \to 0$, which hold for any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$, which satisfy some time scaling relation, $t_\varepsilon / v_\varepsilon \to t \in [0, \infty]$ or $t_\varepsilon / w_\varepsilon \to t \in [0, \infty]$ as $\varepsilon \to 0$, with time scaling factors $v_\varepsilon = p_{\varepsilon,12}^{-1} + p_{\varepsilon,21}^{-1} > w_\varepsilon = (p_{\varepsilon,12} + p_{\varepsilon,21})^{-1} \to \infty$ as $\varepsilon \to 0$. The corresponding limiting probabilities $\pi_{ij}(t, A)$ may depend on parameter t and an initial state i of the modulating semi-Markov process. They take essentially different forms, for cases $t = 0, t \in (0, \infty)$ and $t = \infty$. We classify the corresponding theorems, respectively, as short, long and super-long time individual ergodic theorems.

Individual ergodic theorems for singularly and super-singularly perturbed alternating regenerative processes presented in the paper were not known before.

The main analytic tool used for obtaining ergodic theorems is based on results concerned generalisation of the renewal theorem to the model of perturbed renewal equation given in works [14, 43–45] and quasi-ergodic theorems for perturbed regenerative processes with regenerative lifetimes given in works [13, 14, 38, 53].

Here, works [2, 3, 5, 6, 8, 14, 16, 18–22, 39–42, 56–58, 61, 62] can also be mentioned, where one can find results and bibliographies of works on limit and ergodic type theorems and related problems for singularly perturbed Markov type processes. The difference with some related results presented in these works, is that we operate, in general, with non-Markov regenerative type processes and do not exploit additive accumulation or phase merging phenomena.

We do prefer to use for getting individual ergodic type theorems, as we think, the most effective methods based on generalisations of the classical renewal theorem to model of perturbed renewal equation developed in the above mentioned works [13, 14, 38, 43–45, 53]. This let us get the corresponding ergodic theorems under minimal conditions. In the case of unperturbed and non-alternating regenerative processes, these conditions reduce to the minimal conditions of the classical individual ergodic theorem for unperturbed regenerative processes yielded by the famous renewal theorem, which is given in its final form in [12].

The paper includes 7 sections. In Sect. 3.2, so-called quasi-ergodic theorems for perturbed regenerative processes with regenerative lifetimes, which play the role of basic analytical tool in our studies, and the model of perturbed alternating regenerative processes are presented, and comments concerning regularly, singularly and super-singularly perturbed alternating regenerative processes are given. In Sects. 3.3–3.6, short, long and super-long individual ergodic theorems for regularly, singularly and super-singularly perturbed alternating regenerative processes are presented. Section 3.7, contains a short summary of the results and a list of some directions for future development and improvement of results presented in the paper.

3.2 Perturbed Regenerative and Alternating Regenerative Processes

In this section, we present so-called quasi-ergodic theorems for perturbed regenerative processes with regenerative lifetimes, which play the role of basic analytical tool in our studies, introduce alternating regenerative processes, comment and compare models of regularly, singularly and super-singularly perturbed alternating regenerative processes and forms of the corresponding ergodic theorems, and describe the special procedure of aggregation for regeneration times, which play an important role in ergodic theorems for perturbed alternating regenerative processes.

3.2.1 Quasi-Ergodic Theorems for Perturbed Regenerative Processes with Regenerative Lifetimes

The main tool, which we are going to use are ergodic theorems for perturbed regenerative processes with regenerative lifetimes, given in book [14].

Let $\langle \Omega_\varepsilon, \mathcal{F}_\varepsilon, \mathsf{P}_\varepsilon \rangle$ be, for every $\varepsilon \in [0, 1]$, a probability space. We assume that all stochastic processes and random variables introduced below and indexed by parameter ε are defined on the probability space $\langle \Omega_\varepsilon, \mathcal{F}_\varepsilon, \mathsf{P}_\varepsilon \rangle$.

Let for every $n = 1, 2, \ldots$: **(a)** $\bar{\xi}_{\varepsilon,n} = \langle \xi_{\varepsilon,n}(t), t \geq 0 \rangle$ be a stochastic process with a phase space $\mathbb{X}$ (with the corresponding σ-algebra of measurable subsets $\mathcal{B}_{\mathbb{X}}$), measurable in the sense that $\xi_{\varepsilon,n}(t, \omega)$, $(t, \omega) \in [0, \infty) \times \Omega_\varepsilon$ is measurable function of (t, ω) (this means that $\{(t; \omega) \in A\} \in \mathcal{B}_+ \times \mathcal{F}_\varepsilon$, $A \in \mathcal{B}_{\mathbb{X}}$, where $\mathcal{B}_+ \times \mathcal{F}_\varepsilon$ is the minimal σ-algebra containing all products $B \times C$, $B \in \mathcal{B}_+$, $C \in \mathcal{F}_\varepsilon$, $\mathcal{B}_+$ is the σ-algebra of Borel subsets of $[0, \infty)$); **(b)** $\kappa_{\varepsilon,n}$ be a non-negative random variable; **(c)** $\mu_{\varepsilon,n}$ is a non-negative random variable. Further, we assume that: **(d)** random triplets $\langle \bar{\xi}_{\varepsilon,n} = \langle \xi_{\varepsilon,n}(t), t \geq 0 \rangle, \kappa_{\varepsilon,n}, \mu_{\varepsilon,n} \rangle$, are mutually independent; **(e)** the joint distributions of random variables $\xi_{\varepsilon,n}(t_k)$, $k = 1, \ldots, r$ and $\kappa_{\varepsilon,n}$, $\mu_{\varepsilon,n}$ do not depend on $n \geq 1$, for every $t_k \in [0, \infty)$, $k = 1, \ldots, r$, $r \geq 1$.

Let us define regeneration times, $\tau_{\varepsilon,n} = \kappa_{\varepsilon,1} + \cdots + \kappa_{\varepsilon,n}$, $n = 1, 2, \ldots$, $\tau_{\varepsilon,0} = 0$, a standard regenerative process,

$$\xi_\varepsilon(t) = \xi_{\varepsilon,n}(t - \tau_{\varepsilon,n-1}), \text{ for } t \in [\tau_{\varepsilon,n-1}, \tau_{\varepsilon,n}), \ n = 1, 2, \ldots, \tag{3.3}$$

and a regenerative lifetime,

$$\mu_\varepsilon = \sum_{k=1}^{\nu_\varepsilon - 1} \kappa_{\varepsilon,k} + \mu_{\varepsilon,\nu_\varepsilon} \mathrm{I}(\nu_\varepsilon < \infty), \text{ where } \nu_\varepsilon = \min(n \geq 1 : \mu_{\varepsilon,n} < \kappa_{\varepsilon,n}). \tag{3.4}$$

We exclude instant regenerations and, thus, assume that the following condition holds:

A: $\mathsf{P}\{\kappa_{\varepsilon,1} > 0\} = 1$, for every $\varepsilon \in [0, 1]$.

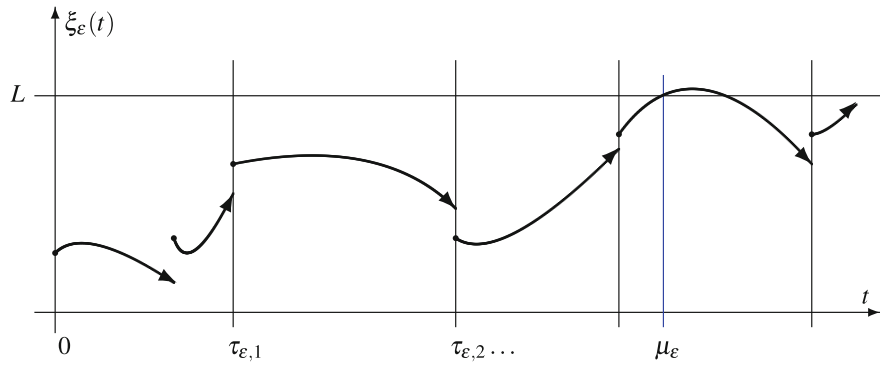

Fig. 3.1 A regenerative process with regenerative lifetime

Condition **A** obviously implies that random variables $\tau_{\varepsilon,n} \xrightarrow{\mathsf{P}} \infty$ as $n \to \infty$, for every $\varepsilon \in [0, 1]$, and, thus, process $\xi_\varepsilon(t)$ is well defined on the time interval $[0, \infty)$.

Figure 3.1 presents an example of trajectory for a real-valued regenerative process $\xi_\varepsilon(t)$ with a regenerative lifetime μ_ε. The role of the regenerative lifetime μ_ε plays, in this case, the first time of exceeding a level L by the regenerative process $\xi_\varepsilon(t)$.

Let us introduce distribution functions $\bar{F}_\varepsilon(t) = \mathsf{P}\{\tau_{\varepsilon,1} \le t\} = \mathsf{P}\{\kappa_{\varepsilon,1} \le t\}, t \ge 0$ and $F_\varepsilon(t) = \mathsf{P}\{\tau_{\varepsilon,1} \le t, \mu_\varepsilon \ge \tau_{\varepsilon,1}\} = \mathsf{P}\{\kappa_{\varepsilon,1} \le t, \mu_{\varepsilon,1} \ge \kappa_{\varepsilon,1}\}, t \ge 0$, and stopping probabilities $f_\varepsilon = 1 - F_\varepsilon(\infty) = \mathsf{P}\{\mu_\varepsilon < \tau_{\varepsilon,1}\} = \mathsf{P}\{\mu_{\varepsilon,1} < \kappa_{\varepsilon,1}\}$. We assume that the following condition holds:

B: **(a)** $\bar{F}_\varepsilon(\cdot) \Rightarrow \bar{F}_0(\cdot)$ as $\varepsilon \to 0$, where $\bar{F}_0(t)$ is a non-arithmetic distribution function, **(b)** $f_\varepsilon \to f_0 = 0$ as $\varepsilon \to 0$.

Here and henceforth symbol $\varepsilon \to 0$ is used to show that $0 < \varepsilon \to 0$.

Condition **B** obviously implies that,

$$F_\varepsilon(\cdot) \Rightarrow F_0(\cdot) \equiv \bar{F}_0(\cdot) \text{ as } \varepsilon \to 0. \tag{3.5}$$

Let us introduce expectations $\bar{e}_\varepsilon = \int_0^\infty s\bar{F}_\varepsilon(ds)$ and $e_\varepsilon = \int_0^\infty sF_\varepsilon(ds)$. We also assume that the following condition holds:

C: **(a)** $\bar{e}_\varepsilon < \infty$, for $\varepsilon \in [0, 1]$, **(b)** $\bar{e}_\varepsilon \to \bar{e}_0$ as $\varepsilon \to 0$.

Condition **B** and **C** obviously imply that $e_\varepsilon < \infty$, for $\varepsilon \in [0, 1]$ and,

$$e_\varepsilon \to e_0 = \bar{e}_0 \text{ as } \varepsilon \to 0. \tag{3.6}$$

The object of our interest are probabilities $P_\varepsilon(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A, \mu_\varepsilon > t\}, A \in \mathcal{B}_{\mathbb{X}}, t \ge 0$. These probabilities are, for every $A \in \mathcal{B}_{\mathbb{X}}$, a measurable function of $t \ge 0$, which is the unique bounded solution for the following renewal equation,

$$P_\varepsilon(t, A) = q_\varepsilon(t, A) + \int_0^t P_\varepsilon(t-s, A) F_\varepsilon(ds),\ t \geq 0, \tag{3.7}$$

where $q_\varepsilon(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A, \tau_{\varepsilon,1} \wedge \mu_\varepsilon > t\} = \mathsf{P}\{\xi_{\varepsilon,1}(t) \in A, \tau_{\varepsilon,1} \wedge \mu_{\varepsilon,1} > t\}$, $A \in \mathcal{B}_{\mathbb{X}}, t \geq 0$.

We also impose the following condition on the functions $q_\varepsilon(t, A)$:

D: There exist a non-empty class of sets $\Gamma \subseteq \mathcal{B}_{\mathbb{X}}$ such that, for every $A \in \Gamma$, the asymptotic relation, $\lim_{u \to 0} \overline{\lim}_{0 \leq \varepsilon \to 0} \sup_{-(u \wedge s) \leq v \leq u} |q_\varepsilon(s+v, A) - q_0(s, A)| = 0$, holds almost everywhere with respect to the Lebesgue measure $m(ds)$ on $[0, \infty)$.

The class Γ appearing in condition **D** contains the phase space $\mathbb{X}$ and is closed with respect to the operation of union for not intersecting sets, the operation of difference for sets connected by relation of inclusion, and the complement operation. The detailed comments are given in Sect. 3.2.5.

Conditions **A**–**D** imply that process $\xi_0(t), t \geq 0$ is ergodic and the following asymptotic relation holds, for $A \in \Gamma$,

$$P_0(t, A) \to \pi_0(A) \text{ as } t \to \infty, \tag{3.8}$$

where $\pi_0(A)$ is the corresponding stationary distribution given by the following relation,

$$\pi_0(A) = \frac{1}{e_0} \int_0^\infty q_0(s, A) m(ds),\ A \in \mathcal{B}_{\mathbb{X}}. \tag{3.9}$$

Now we are prepared to formulate a theorem, which is a variant of the quasi-ergodic theorem, for the perturbed regenerative processes with regenerative lifetimes given in book [14]. It is also worth to note that this theorem is the direct corollary of the version renewal theorem for perturbed renewal equation given in papers [43–45].

Theorem 3.1 *Let conditions* **A**–**D** *hold. Then, for every* $A \in \Gamma$, *and any* $0 \leq t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $f_\varepsilon t_\varepsilon \to t \in [0, \infty]$ *as* $\varepsilon \to 0$,

$$P_\varepsilon(t_\varepsilon, A) \to e^{-t/e_0} \pi_0(A) \text{ as } \varepsilon \to 0. \tag{3.10}$$

Let us now assume that the model assumption **(e)** formulated above holds only for $n \geq 2$. In this case, the process $\xi_\varepsilon(t), t \geq 0$ is usually referred as a regenerative process with transition period $[0, \tau_{\varepsilon,1})$.

We also shall use the extension of Theorem 3.1 on the model of perturbed regenerative processes with transition period. In this case, the shifted process $\xi_\varepsilon^{(1)}(t) = \xi_\varepsilon(\tau_{\varepsilon,1} + t), t \geq 0$ is a standard regenerative process, with regeneration times $\tau_{\varepsilon,n}^{(1)} = \kappa_{\varepsilon,2} + \cdots + \kappa_{\varepsilon,n+1}, n = 1, 2, \ldots, \tau_{\varepsilon,0}^{(1)} = 0$ and the corresponding shifted regenerative lifetime $\mu_\varepsilon^{(1)} = \sum_{k=1}^{\nu_\varepsilon^{(1)}-1} \kappa_{\varepsilon,1+k} + \mu_{\varepsilon,1+\nu_\varepsilon^{(1)}} \mathrm{I}(\nu_\varepsilon^{(1)} < \infty)$, where $\nu_\varepsilon^{(1)} = \min(n \geq 1 : \mu_{\varepsilon,1+n} < \kappa_{\varepsilon,1+n})$.

All quantities appearing in conditions **A**–**D**, the renewal equation (3.7) and relation (3.9) should be defined using shifted sequence of triplets $\langle \bar{\xi}_{\varepsilon,2} = \langle \xi_{\varepsilon,2}(t), t \geq 0\rangle, \kappa_{\varepsilon,2}, \mu_{\varepsilon,2}\rangle$. It is also natural to index the above mentioned quantities by the upper index $^{(1)}$, for example, to use notation $P_\varepsilon^{(1)}(t, A) = \mathsf{P}\{\xi_\varepsilon^{(1)}(t) \in A, \mu_\varepsilon^{(1)} > t\}$, etc. Probabilities $P_\varepsilon^{(1)}(t, A)$ satisfy the renewal equation (3.7). Theorem 3.1 presents, in this case, the corresponding ergodic relation for these probabilities.

Probabilities $P_\varepsilon(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A, \mu_\varepsilon > t\}$, defined for the initial regenerative process with transition period, are, for every $A \in \mathcal{B}_{\mathbb{X}}$, connected with probabilities $P_\varepsilon^{(1)}(t_\varepsilon, A)$ by the following renewal type transition relation,

$$P_\varepsilon(t, A) = \tilde{q}_\varepsilon(t, A) + \int_0^t P_\varepsilon^{(1)}(t - s, A)\tilde{F}_\varepsilon(ds), \ t \geq 0, \tag{3.11}$$

where $\tilde{q}_\varepsilon(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A, \tau_{\varepsilon,1} \wedge \mu_\varepsilon > t\} = \mathsf{P}\{\xi_{\varepsilon,1}(t) \in A, \tau_{\varepsilon,1} \wedge \mu_{\varepsilon,1} > t\}$, $A \in \mathcal{B}_{\mathbb{X}}, t \geq 0$ and $\tilde{F}_\varepsilon(t) = \mathsf{P}\{\tau_{\varepsilon,1} \leq t, \mu_{\varepsilon,1} \geq \tau_{\varepsilon,1}\}, t \geq 0$ are the corresponding characteristics related to the transition period.

We admit that the transition period can be of zero duration and, thus, the distribution function $\tilde{F}_\varepsilon(t)$ can possess an atom in zero or even be concentrated at zero, for $\varepsilon \in [0, 1]$.

Let us additionally assume that the following condition holds:

E: $\tilde{F}_\varepsilon(\cdot) \Rightarrow \tilde{F}_0(\cdot)$ as $\varepsilon \to 0$, where $\tilde{F}_0(t)$ is a proper distribution function.

Let also $\tilde{f}_\varepsilon = \mathsf{P}\{\mu_{\varepsilon,1} < \tau_{\varepsilon,1}\} = 1 - \tilde{F}_\varepsilon(\infty)$. Condition **E** obviously implies that the stopping probabilities for transition period, $\tilde{f}_\varepsilon \to \tilde{f}_0 = 0$ as $\varepsilon \to 0$.

It is also useful to note that $\tilde{q}_\varepsilon(t, A) \leq \mathsf{P}\{\tau_{\varepsilon,1} \wedge \mu_{\varepsilon,1} > t\} = \mathsf{P}\{\tau_{\varepsilon,1} > t, \mu_{\varepsilon,1} \geq \tau_{\varepsilon,1}\} + \mathsf{P}\{\tau_{\varepsilon,1} \wedge \mu_{\varepsilon,1} > t, \mu_{\varepsilon,1} < \tau_{\varepsilon,1}\} \leq \mathsf{P}\{\mu_{\varepsilon,1} \geq \tau_{\varepsilon,1}\} - \mathsf{P}\{\tau_{\varepsilon,1} \leq t, \mu_{\varepsilon,1} \geq \tau_{\varepsilon,1}\} + \mathsf{P}\{\mu_{\varepsilon,1} < \tau_{\varepsilon,1}\} = \tilde{F}_\varepsilon(\infty) - \tilde{F}_\varepsilon(t) + \tilde{f}_\varepsilon$. This relation and condition **E** imply that $\tilde{q}_\varepsilon(t_\varepsilon, A) \to 0$ as $\varepsilon \to 0$, for any $0 \leq t_\varepsilon \to \infty$ as $\varepsilon \to 0$.

The following quasi-ergodic theorem for perturbed regenerative processes with transition period, is also given in book [14].

Theorem 3.2 *Let conditions* **A**–**E** *hold. Then, for every* $A \in \Gamma$, *and any* $0 \leq t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $f_\varepsilon t_\varepsilon \to t \in [0, \infty]$ *as* $\varepsilon \to 0$,

$$P_\varepsilon(t_\varepsilon, A) \to e^{-t/e_0}\pi_0(A) \text{ as } \varepsilon \to 0. \tag{3.12}$$

In the case of standard regenerative processes, Theorem 3.2 just reduces to Theorem 3.1. Indeed, condition **E** can be omitted since it is implied by condition **B**. The ergodic relation (3.12) reduces to the ergodic relation (3.10).

Let us also introduce modified regenerative lifetimes $\mu_{\varepsilon,-} = \sum_{k=1}^{\nu_\varepsilon - 1} \kappa_{\varepsilon,k}$, and $\mu_{\varepsilon,+} = \sum_{k=1}^{\nu_\varepsilon} \kappa_{\varepsilon,k}$ and consider probabilities $P_{\varepsilon,\pm}(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A, \mu_{\varepsilon,\pm} > t\}$, $A \in \mathcal{B}_{\mathbb{X}}, t \geq 0$.

Obviously, $\mu_{\varepsilon,-} \leq \mu_\varepsilon \leq \mu_{\varepsilon,+}$ and, thus, $P_{\varepsilon,-}(t, A) \leq P_\varepsilon(t, A) \leq P_{\varepsilon,+}(t, A)$, for any $A \in \mathcal{B}_{\mathbb{X}}, t \geq 0$.

The following theorem is a useful modification of Theorem 3.2.

Theorem 3.3 *Let conditions* **A**–**E** *hold. Then, for every* $A \in \Gamma$, *and any* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $f_\varepsilon t_\varepsilon \to t \in [0, \infty]$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,\pm}(t_\varepsilon, A) \to e^{-t/e_0}\pi_0(A) \text{ as } \varepsilon \to 0. \tag{3.13}$$

Proof Conditions **A**–**C** imply that the asymptotic relation, $f_\varepsilon \kappa_{\varepsilon,\nu_\varepsilon} \mathrm{I}(\nu_\varepsilon < \infty) \xrightarrow{P} 0$ as $\varepsilon \to 0$, holds. The asymptotic relation (3.13) is an obvious corollary of this asymptotic relation and the ergodic relation (3.12) given in Theorem 3.2. □

3.2.2 One-Dimensional and Multi-dimensional Distributions for Perturbed Regenerative Processes

Individual ergodic theorems formulated in Theorems 3.1–3.3 present ergodic relations for one-dimensional distributions $P_\varepsilon(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A, \mu_\varepsilon > t\}$ for regenerative processes with regenerative lifetimes.

It possible to weaken the model assumption **(e)** formulated in Sect. 3.2.1. This assumption concerns multi-dimensional joint distributions of random variables $\xi_{\varepsilon,n}(t_k), k = 1, \ldots, r, \kappa_{\varepsilon,n}$ and $\mu_{\varepsilon,n}$.

It can be replaced by the weaker assumption that the joint distributions of random variables $\xi_{\varepsilon,n}(t), \kappa_{\varepsilon,n}$ and $\mu_{\varepsilon,n}$ do not depend on $n \ge 1$, for every $t \ge 0$.

The process $\xi_\varepsilon(t), t \ge 0$ will still process the corresponding weaken, say, one-dimensional regenerative property, which, in fact, means that one-dimensional distributions $P_\varepsilon(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A, \mu_\varepsilon > t\}, t \ge 0$ satisfy the renewal equation (3.7).

Formulations of conditions **A**–**E** as well as propositions of Theorems 3.1–3.3 still remain to be valid.

3.2.3 Ergodic Theorems for Standard Regenerative Processes

We would like to mention the important case, where stopping probability $f_\varepsilon = 0, \varepsilon \in [0, 1]$. In this case, the regenerative stopping time $\mu_\varepsilon = \infty$ with probability 1. Also, $f_\varepsilon t_\varepsilon \to 0$ as $\varepsilon \to 0$, for any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$.

Probability $P_\varepsilon(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A\}$ is a one-dimensional distribution for process $\xi_\varepsilon(t)$. Theorems 3.1–3.3 present in this case usual individual ergodic theorems for perturbed regenerative processes $\xi_\varepsilon(t)$.

It is also worth to mention the case of unperturbed regenerative process $\xi_0(t), t \ge 0$. Conditions **A**–**D** reduce in this case to the minimal conditions of the individual ergodic theorem for regenerative processes, which directly follows from the renewal theorem given in its final form in [12]: (a) $F_0(\cdot)$ is a non-arithmetic distribution function without an atom in zero; (b) $e_0 = \int_0^\infty s F_0(ds) < \infty$; (c) function $q_0(s, A), s \ge 0$ is, for $A \in \Gamma$, continuous almost everywhere with respect to the Lebesgue measure $m(ds)$ on $[0, \infty)$.

Note that $q_0(s, A) \le 1 - F_0(s)$, $s \ge 0$ and, thus, under the above condition (b), condition (c) is equivalent to the assumption of direct Riemann integrability of the free term in the renewal equation (3.7), imposed on this term in the renewal theorem given in [12].

Also, condition **E** just reduces to the assumption that (d) $\tilde{F}_0(\cdot)$ is a proper distribution function.

The corresponding individual ergodic theorem takes in this case the form of the asymptotic relation (3.8), i.e., $P_0(t, A) \to \pi_0(A)$ as $t \to \infty$, for $A \in \Gamma$.

3.2.4 Perturbed Alternating Regenerative Processes

Let, for every $i = 1, 2$ and $n = 1, 2, \ldots$: **(f)** $\bar{\xi}_{\varepsilon,i,n} = \langle \xi_{\varepsilon,i,n}(t), t \ge 0 \rangle$ be a measurable stochastic process with a phase space $\mathbb{X}$; **(g)** $\kappa_{\varepsilon,i,n}$ be a non-negative random variable; **(h)** $\eta_{\varepsilon,i,n}$ and η_ε are binary random variables taking values in the space $\mathbb{Y} = \{1, 2\}$. Further, we assume that: **(i)** triplets $\langle \bar{\xi}_{\varepsilon,i,n} = \langle \xi_{\varepsilon,i,n}(t), t \ge 0 \rangle, \kappa_{\varepsilon,i,n}, \eta_{\varepsilon,i,n} \rangle$, $i = 1, 2, n = 1, 2, \ldots$ and the random variable η_ε are mutually independent; **(j)** the joint distributions of random variables $\xi_{\varepsilon,i,n}(t_k)$, $k = 1, \ldots, r$ and $\kappa_{\varepsilon,i,n}$, $\eta_{\varepsilon,i,n}$ do not depend on $n \ge 1$, for every $i = 1, 2$ and $t_k \in [0, \infty)$, $k = 1, \ldots, r$, $r \ge 1$.

Here, the measurability assumption for processes $\bar{\xi}_{\varepsilon,i,n}$ is absolutely analogous to those formulated in the model assumption **(a)** for processes $\bar{\xi}_{\varepsilon,n}$.

Let us define recurrently stochastic sequences of switching binary random indices $\eta_{\varepsilon,n}$, $n = 0, 1, \ldots$ and regeneration times $\tau_{\varepsilon,n}$, $n = 0, 1, \ldots$ by the following recurrent relations, $\eta_{\varepsilon,n} = \eta_{\varepsilon,\eta_{\varepsilon,n-1},n}$, $n = 1, 2, \ldots$, $\eta_{\varepsilon,0} = \eta_\varepsilon$ and $\tau_{\varepsilon,n} = \kappa_{\varepsilon,\eta_{\varepsilon,0},1} + \cdots + \kappa_{\varepsilon,\eta_{\varepsilon,n-1},n}$, $n = 1, 2, \ldots$, $\tau_{\varepsilon,0} = 0$, and the modulated alternating regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \ge 0$ by the following recurrent relations,

$$\xi_\varepsilon(t) = \xi_{\varepsilon,\eta_{\varepsilon,n-1},n}(t - \tau_{\varepsilon,n-1}) \text{ and } \eta_\varepsilon(t) = \eta_{\varepsilon,n-1},$$
$$\text{for } t \in [\tau_{\varepsilon,n-1}, \tau_{\varepsilon,n}), \quad n = 1, 2, \ldots. \tag{3.14}$$

Figure 3.2 given below presents an example of trajectory of an alternating regenerative process $\xi_\varepsilon(t)$ and the corresponding modulating semi-Markov index process $\eta_\varepsilon(t)$.

We exclude instant regenerations and, thus, assume that the following condition holds:

F: $\mathsf{P}\{\kappa_{\varepsilon,i,1} > 0\} = 1$, $i = 1, 2$, for every $\varepsilon \in [0, 1]$.

This condition obviously implies that $\tau_{\varepsilon,n} \xrightarrow{\mathsf{P}} \infty$ as $n \to \infty$, for every $\varepsilon \in [0, 1]$, and thus, the above alternating regenerative process is well defined on the time interval $[0, \infty)$.

Now, let us formulate conditions, which make it possible to consider $(\xi_0(t), \eta_0(t))$, $t \ge 0$ as an unperturbed process and $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \ge 0$ as its perturbed version, for $\varepsilon \in (0, 1]$.

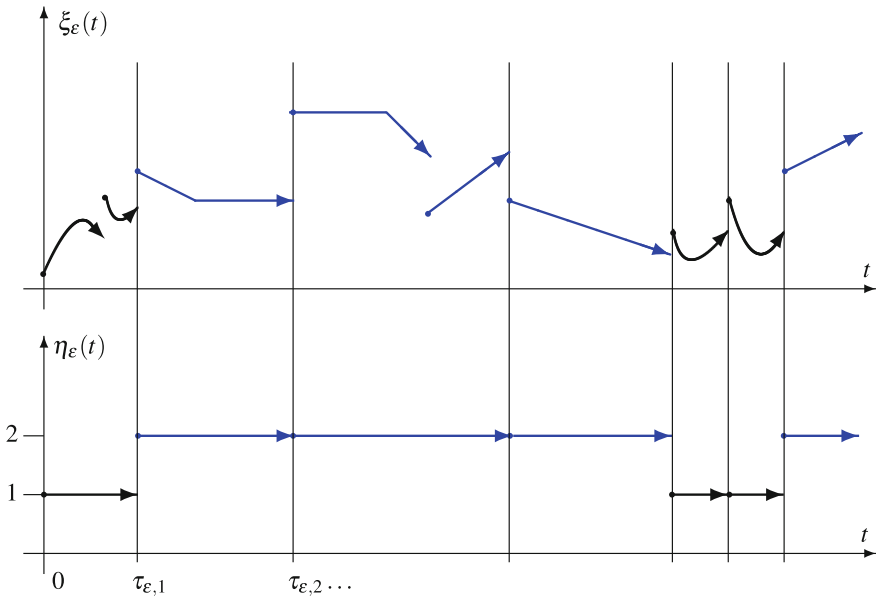

Fig. 3.2 A regenerative process with regenerative lifetime

The above model assumptions **(f)**–**(j)** imply that the modulating index sequence $\eta_{\varepsilon,n}$, $n = 0, 1, \ldots$ is a homogeneous Markov chain with the phase space $\mathbb{Y} = \{1, 2\}$, the initial distribution $\bar{p}_\varepsilon = \langle p_{\varepsilon,i} = \mathsf{P}\{\eta_{\varepsilon,0} = i\}, i = 1, 2\rangle$, and transition probabilities, $p_{\varepsilon,ij} = \mathsf{P}\{\eta_{\varepsilon,1} = j/\eta_{\varepsilon,0} = i\} = \mathsf{P}\{\eta_{\varepsilon,i,1} = j\}$, $i, j = 1, 2$. We assume that the following condition holds:

G: **(a)** $p_{\varepsilon,ij} = 0, \varepsilon \in (0, 1]$ or $p_{\varepsilon,ij} > 0, \varepsilon \in (0, 1]$, for $i, j = 1, 2$;
(b) $p_{\varepsilon,ij} \to p_{0,ij}$ as $\varepsilon \to 0$, for $i, j = 1, 2$.

The above model assumptions **(f)**–**(j)** also imply that the modulating index process $\eta_\varepsilon(t), t \geq 0$ is a semi-Markov process with the phase space $\mathbb{Y}$ and transition probabilities, $Q_{\varepsilon,ij}(t) = \mathsf{P}\{\tau_{\varepsilon,1} \leq t, \eta_{\varepsilon,1} = j/\eta_{\varepsilon,0} = i\} = \mathsf{P}\{\kappa_{\varepsilon,i,1} \leq t, \eta_{\varepsilon,i,1} = j\}$, $t \geq 0, i, j = 1, 2$. Also, let us introduce conditional distribution functions $F_{\varepsilon,ij}(t) = Q_{\varepsilon,ij}(t)/p_{\varepsilon,ij}, t \geq 0$ defined for $i, j \in \mathbb{Y}$ such that $p_{\varepsilon,ij} > 0, \varepsilon \in (0, 1]$.

We also assume that the following condition holds:

H: **(a)** $Q_{\varepsilon,ij}(\cdot) \Rightarrow Q_{0,ij}(\cdot)$ as $\varepsilon \to 0$, for $i, j = 1, 2$,
(b) $Q_{0,ij}(t) = 0, t \geq 0$ if $p_{0,ij} = 0$ or $F_{0,ij}(t) = Q_{0,ij}(t)/p_{0,ij}, t \geq 0$ is a non-arithmetic distribution function if $p_{0,ij} > 0$.

Remark 3.1 Conditions of convergence **G (b)** and **H (a)** can be reformulated in terms of Laplace transforms $\phi_{\varepsilon,ij}(s) = \int_0^\infty e^{-st} Q_{\varepsilon,ij}(dt)$, $s \geq 0, i, j = 1, 2$. These conditions are equivalent to the assumption that $\phi_{\varepsilon,ij}(s) \to \phi_{0,ij}(s)$ as $\varepsilon \to 0$, for $s \geq 0$ and $i, j = 1, 2$.

Let us introduce expectations, $e_{\varepsilon,ij} = \mathsf{E}_i \tau_{\varepsilon,1} \mathrm{I}(\eta_{\varepsilon,1} = j) = \mathsf{E}\kappa_{\varepsilon,i,1} \mathrm{I}(\eta_{\varepsilon,i,1} = j) = \int_0^\infty s\, Q_{\varepsilon,ij}(ds)$, $i, j = 1, 2$ and $e_{\varepsilon,i} = \mathsf{E}_i \tau_{\varepsilon,1} = \mathsf{E}\kappa_{\varepsilon,i,1} = e_{\varepsilon,i1} + e_{\varepsilon,i2}$, $i = 1, 2$.

Here and henceforth, we use notations P_i and E_i for conditional probabilities and expectations under condition $\eta_\varepsilon(0) = \eta_\varepsilon = i$.

We also impose the following condition of convergence for above expectations:

I: **(a)** $e_{\varepsilon,ij} < \infty$, for every $\varepsilon \in [0, 1]$ and $i = 1, 2$;
(b) $e_{\varepsilon,ij} \to e_{0,ij}$ as $\varepsilon \to 0$, for $i = 1, 2$.

The object of our interest is the joint distributions,

$$P_{\varepsilon,ij}(t, A) = \mathsf{P}_i\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = j\}, A \in \mathcal{B}_{\mathbb{X}}, i, j = 1, 2, t \geq 0. \tag{3.15}$$

Probabilities $P_{\varepsilon,ij}(t, A)$ are, for every $A \in \mathcal{B}_{\mathbb{X}}$, $j = 1, 2$, measurable functions of $t \geq 0$, which are the unique bounded solution for the following system of renewal type equations,

$$P_{\varepsilon,ij}(t, A) = \delta(i, j) q_{\varepsilon,i}(t, A) + \sum_{k=1}^{2} \int_0^t P_{\varepsilon,kj}(t - s, A) Q_{\varepsilon,ik}(ds),\ t \geq 0, i = 1, 2. \tag{3.16}$$

where $q_{\varepsilon,i}(t, A) = \mathsf{P}_i\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = i, \tau_{\varepsilon,1} > t\} = \mathsf{P}\{\xi_{\varepsilon,i,1}(t) \in A, \kappa_{\varepsilon,i,1} > t\}$, $A \in \mathcal{B}_{\mathbb{X}}$, $t \geq 0, i, j = 1, 2$.

Finally, we also impose the following condition on functions $q_{\varepsilon,i}(t, A)$:

J: There exists a non-empty class of sets $\Gamma \subseteq \mathcal{B}_{\mathbb{X}}$ such that, for every $A \in \Gamma$, asymptotic relations, $\lim_{u \to 0} \overline{\lim}_{0 \leq \varepsilon \to 0} \sup_{-(u \wedge s) \leq v \leq u} |q_{\varepsilon,i}(s + v, A) - q_{0,i}(s, A)| = 0$, $i = 1, 2$, hold almost everywhere with respect to the Lebesgue measure $m(ds)$ on $[0, \infty)$.

As for condition **D**, the class Γ appearing in condition **J** contains the phase space $\mathbb{X}$ and is closed with respect to the operation of union for not intersecting sets, the operation of difference for sets connected by relation of inclusion, and the complement operation. The corresponding comments are given below, in Sect. 3.2.5.

Consider also, for $i = 1, 2$, the standard regenerative process $\xi_{\varepsilon,i}(t)$, $t \geq 0$ with regeneration times $\tau_{\varepsilon,i,n} = \kappa_{\varepsilon,i,1} + \cdots + \kappa_{\varepsilon,i,n}$, $n = 1, 2, \ldots$, $\tau_{\varepsilon,i,0} = 0$, defined by the recurrent relations, $\xi_{\varepsilon,i}(t) = \xi_{\varepsilon,i,n}(t - \tau_{\varepsilon,i,n-1})$, for $t \in [\tau_{\varepsilon,i,n-1}, \tau_{\varepsilon,in})$, $n = 1, 2, \ldots$.

Conditions **F**–**J** imply that, for every $i = 1, 2$, all conditions of Theorem 3.1 hold for regenerative process $\xi_{\varepsilon,i}(t)$, with the corresponding stopping probabilities $f_{\varepsilon,i} = 0, \varepsilon \in [0, 1]$.

Thus, for every $i = 1, 2$, the following ergodic relation holds, for $A \in \Gamma$ and any $0 \leq t_\varepsilon \to \infty$ as $\varepsilon \to 0$,

$$\mathsf{P}\{\xi_{\varepsilon,i}(t_\varepsilon) \in A\} \to \pi_{0,i}(A) \text{ as } \varepsilon \to 0, \tag{3.17}$$

where the probabilities $\pi_{0,i}(A)$ are corresponding stationary probabilities for the regenerative process $\xi_{0,i}(t)$ given by the following relation,

$$\pi_{0,i}(A) = \frac{1}{e_{0,i}} \int_0^\infty q_{0,i}(s, A)m(ds), \ A \in \mathcal{B}_{\mathbb{X}}. \tag{3.18}$$

3.2.5 *Stricture of Class* Γ

Note that functions $q_\varepsilon(s, A)$ and $q_{\varepsilon,i}(s, A)$ appearing, respectively, in conditions **D** and **J** are finite measures as functions of $A \in \mathcal{B}_{\mathbb{X}}$.

This, in obvious way, implies that the class Γ appearing in condition **D** or **J** is closed with respect to the operation of union for non-intersecting sets, i.e., if the convergence relation given in condition **D** or **J** holds for sets A' and A'' such that $A' \cap A'' = \emptyset$, then this relation also holds for set $A = A' \cup A''$.

The class Γ appearing in condition **D** or **J** also is closed with respect to the operation of differences for sets connected by relation of inclusion, i.e., if the convergence relation given in condition **D** or **J** holds for sets A' and A'' such that $A' \subseteq A''$, then this relation also holds for set $A = A'' \setminus A'$.

Also, the class of sets Γ appearing in condition **D** or **J** includes the phase space $\mathbb{X}$ under assumption that, respectively, condition **B** holds or conditions **G** and **H** hold.

Let us check this, for example, for the case of condition **D**. Indeed,

$$\begin{aligned} q_\varepsilon(t, \mathbb{X}) &= \mathsf{P}\{\tau_{\varepsilon,1} \wedge \mu_{\varepsilon,1} > t\} \\ &= \mathsf{P}\{\tau_{\varepsilon,1} > t, \mu_{\varepsilon,1} \geq \tau_{\varepsilon,1}\} + \mathsf{P}\{\tau_{\varepsilon,1} \wedge \mu_{\varepsilon,1} > t, \mu_{\varepsilon,1} < \tau_{\varepsilon,1}\}. \end{aligned} \tag{3.19}$$

and

$$\begin{aligned} \mathsf{P}\{\tau_{\varepsilon,1} > t, \mu_{\varepsilon,1} \geq \tau_{\varepsilon,1}\} &= \mathsf{P}\{\mu_{\varepsilon,1} \geq \tau_{\varepsilon,1}\} - \mathsf{P}\{\tau_{\varepsilon,1} \leq t, \mu_{\varepsilon,1} \geq \tau_{\varepsilon,1}\} \\ &= F_\varepsilon(\infty) - F_\varepsilon(t). \end{aligned} \tag{3.20}$$

Condition **B** implies that $F_\varepsilon(\infty) - F_\varepsilon(t_\varepsilon) \to 1 - F_0(t)$ as $\varepsilon \to 0$, for any $t_\varepsilon \to t$ as $\varepsilon \to 0$ and $t \in \mathbb{C}(F_0)$, where $\mathbb{C}(F_0)$ is the set of continuity points for the distribution function $F_0(\cdot)$. Also, $\mathsf{P}\{\tau_{\varepsilon,1} \wedge \mu_{\varepsilon,1} > t_\varepsilon, \mu_{\varepsilon,1} < \tau_{\varepsilon,1}\} \leq \mathsf{P}\{\mu_{\varepsilon,1} < \tau_{\varepsilon,1}\} = f_\varepsilon \to 0$ as $\varepsilon \to 0$. The above relations imply that $q_\varepsilon(t_\varepsilon, \mathbb{X}) \to q_0(t, \mathbb{X}) = 1 - F_0(t)$ as $\varepsilon \to 0$, for any $t_\varepsilon \to t \in \mathbb{C}(F_0)$ as $\varepsilon \to 0$. Since $\overline{\mathbb{C}}(F_0) = [0, \infty) \setminus \mathbb{C}(F_0)$ is at most a countable set, $m(\overline{\mathbb{C}}(F_0)) = 0$. The above relations imply, by Lemma 3.2 given in Sect. 3.4.3, that the asymptotic relation appearing in condition **D** holds for $\Gamma = \mathbb{X}$.

Finally, the above remarks imply that class Γ appearing in condition **D** or **J** is also closed with respect to the complement operation, i.e., if the convergence relation given in condition **D** holds for set A, then it also holds for set $\overline{A}$.

Let us, for example, consider the model, where the phase space $\mathbb{X} = \{1, 2, \ldots, m\}$ is a finite set and $\mathcal{B}_{\mathbb{X}}$ is the σ-algebra of all subsets of $\mathbb{X}$. In this case, it is natural

to assume that the corresponding locally uniform convergence relation appearing in condition **J** holds for all one-point sets $A = \{j\}$, $j \in \mathbb{X}$. This will obviously imply that this convergence relation also holds for any subset $A \subseteq \mathbb{X}$ that means that, in this case, class $\Gamma = \mathcal{B}_{\mathbb{X}}$.

3.2.6 Regularly, Singularly and Super-Singularly Perturbed Alternating Regenerative Processes

The aim of the present paper is to give a detailed analysis of individual ergodic theorems for probabilities $P_{\varepsilon,ij}(t, A)$ that is to describe possible variants of their asymptotic behaviour as $t \to \infty$ and $\varepsilon \to 0$.

We shall see that the asymptotic behaviour of transition probabilities $p_{\varepsilon,ij}$, $i, j = 1, 2$ for the Markov chains $\eta_{\varepsilon,n}$ plays an important role in these ergodic theorems. Note that, according to condition **G**, these transition probabilities converge to the corresponding transition probabilities $p_{0,ij}$, $i, j = 1, 2$ for the Markov chain $\eta_{0,n}$, as $\varepsilon \to 0$.

There are three classes of perturbed alternating regenerative processes, with essentially different ergodic properties.

The first class includes so-called "regularly" perturbed alternating regenerative processes, for which the limiting Markov chain $\eta_{0,n}$ is ergodic that, in this case, is equivalent to the assumption that at least one of its transition probabilities $p_{0,12}$ and $p_{0,21}$ is positive.

Here, parameter $\beta = p_{0,12}/p_{0,21}$ plays the key role. Obviously, **(a)** $\beta \in (0, \infty)$, if $p_{0,12}, p_{0,21} > 0$, **(b)** $\beta = 0$, if $p_{0,12} = 0$, $p_{0,21} > 0$, and **(c)** we should count $\beta = \infty$, if $p_{0,12} > 0$, $p_{0,21} = 0$. In case **(a)**, the phase space $\mathbb{Y}$ is one class of communicative states and the corresponding stationary probabilities $\alpha_1(\beta) = \frac{p_{0,21}}{p_{0,12}+p_{0,21}} = \frac{1}{1+\beta}$ and $\alpha_2(\beta) = \frac{p_{0,12}}{p_{0,12}+p_{0,21}} = \frac{1}{1+\beta^{-1}}$. In case **(b)**, the phase space $\mathbb{Y}$ consists of the absorbing state 1 and the transient state 2. In this case, $\alpha_1(0) = 1$ and $\alpha_2(0) = 0$. Analogously, in case **(c)**, the phase space $\mathbb{Y}$ consists of the absorbing state 2 and the transient state 1. In this case $\alpha_1(\infty) = 0$ and $\alpha_2(\infty) = 1$.

In ergodic theorems for perturbed alternating regenerative processes, the asymptotic stability of stationary probabilities for Markov chains $\eta_{\varepsilon,n}$ play the key role. In the case of regularly perturbed models, condition **G** obviously implies that the Markov chain $\eta_{\varepsilon,n}$ is ergodic, for every $\varepsilon \in [0, 1]$. Its stationary probabilities are determined by parameter $\beta_\varepsilon = p_{\varepsilon,12}/p_{\varepsilon,21}$, namely, $\alpha_1(\beta_\varepsilon) = \frac{p_{\varepsilon,21}}{p_{\varepsilon,12}+p_{\varepsilon,21}} = \frac{1}{1+\beta_\varepsilon}$ and $\alpha_2(\beta_\varepsilon) = \frac{p_{\varepsilon,12}}{p_{\varepsilon,12}+p_{\varepsilon,21}} = \frac{1}{1+\beta_\varepsilon^{-1}}$. Condition **G** implies that $\beta_\varepsilon \to \beta$ as $\varepsilon \to 0$ and, in sequel, $\alpha_1(\beta_\varepsilon) \to \alpha_1(\beta)$ and $\alpha_2(\beta_\varepsilon) \to \alpha_2(\beta)$ as $\varepsilon \to 0$.

We shall see that individual ergodic theorems for regularly perturbed alternating processes have a form of asymptotic relation, $P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,j}^{(\beta)}(A)$ as $\varepsilon \to 0$, which holds for $A \in \Gamma$, $i, j = 1, 2$ and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$,

The limiting probabilities $\pi_{0,j}^{(\beta)}(A)$ depend on parameter $\beta \in [0, \infty]$, but they do not depend on an initial state $i \in \mathbb{Y}$. The forms of ergodic theorems are analogous to those, which are known for unperturbed alternating regenerative processes.

The second and the third classes include so-called "singularly" and "super-singularly" perturbed alternating regenerative processes, for which the limiting Markov chain $\eta_{0,n}$ is not ergodic that is equivalent to the assumption that both transition probabilities $p_{0,12}$ and $p_{0,21}$ equal 0.

According condition **G**, four cases are possible. The case **(d)** $0 < p_{\varepsilon,12},\ p_{\varepsilon,21} \to 0$ as $\varepsilon \to 0$, corresponds to singularly perturbed alternating regenerative processes. Three cases, where **(e)** $p_{\varepsilon,12} = 0, \varepsilon \in [0, 1]$ and $0 < p_{\varepsilon,21} \to 0$ as $\varepsilon \to 0$, or **(f)** $0 < p_{\varepsilon,12} \to 0$ as $\varepsilon \to 0$ and $p_{\varepsilon,21} = 0, \varepsilon \in [0, 1]$, or **(g)** $p_{\varepsilon,12},\ p_{\varepsilon,21} = 0, \varepsilon \in [0, 1]$, correspond to super-singularly perturbed alternating regenerative processes.

In case **(d)**, the asymptotic stability for stationary probabilities $\alpha_j(\beta_\varepsilon),\ j = 1, 2$ is provided by the following additional balancing condition that should be assumed to hold for some $\beta \in [0, \infty]$:

$\mathbf{K}_\beta$: $\beta_\varepsilon = p_{\varepsilon,12}/p_{\varepsilon,21} \to \beta$ as $\varepsilon \to 0$.

Condition **G** implies that the Markov chain $\eta_{\varepsilon,n}$ is ergodic, for $\varepsilon \in (0, 1]$. Its stationary probabilities are determined by parameter $\beta_\varepsilon = p_{\varepsilon,12}/p_{\varepsilon,21}$, namely, $\alpha_1(\beta_\varepsilon) = \frac{p_{\varepsilon,21}}{p_{\varepsilon,12}+p_{\varepsilon,21}} = \frac{1}{1+\beta_\varepsilon}$ and $\alpha_2(\beta_\varepsilon) = \frac{p_{\varepsilon,12}}{p_{\varepsilon,12}+p_{\varepsilon,21}} = \frac{1}{1+\beta_\varepsilon^{-1}}$. Conditions **G** and $\mathbf{K}_\beta$ imply that $\beta_\varepsilon \to \beta$ and, in sequel, $\alpha_1(\beta_\varepsilon) \to \alpha_1(\beta) = \frac{1}{1+\beta}$ and $\alpha_2(\beta_\varepsilon) \to \alpha_2(\beta) = \frac{1}{1+\beta^{-1}}$ as $\varepsilon \to 0$.

In case **(e)**, $\beta_\varepsilon = p_{\varepsilon,12}/p_{\varepsilon,21} = 0$, for $\varepsilon \in [0, 1]$, and, thus, condition $\mathbf{K}_0$ holds. Condition **G** implies that the Markov chain $\eta_{\varepsilon,n}$ is ergodic, for $\varepsilon \in (0, 1]$ and its stationary probabilities $\alpha_{\varepsilon,1}(0) = 1, \alpha_{\varepsilon,2}(0) = 0$, for $\varepsilon \in (0, 1]$. Obviously, relations $\alpha_{\varepsilon,1}(0) \to \alpha_1(0) = 1$ and $\alpha_{\varepsilon,2}(0) \to \alpha_2(0) = 0$ as $\varepsilon \to 0$ also hold. Analogously, in the case **(f)**, $\beta_\varepsilon = p_{\varepsilon,12}/p_{\varepsilon,21} = \infty$, for $\varepsilon \in [0, 1]$, and, thus, condition $\mathbf{K}_\infty$ holds. Condition **G** implies that the Markov chain $\eta_{\varepsilon,n}$ is ergodic, for $\varepsilon \in (0, 1]$ and its stationary probabilities $\alpha_{\varepsilon,1}(\infty) = 0, \alpha_{\varepsilon,2}(\infty) = 1$, for $\varepsilon \in (0, 1]$. Obviously, relations $\alpha_{\varepsilon,1}(\infty) \to \alpha_1(\infty) = 0$ and $\alpha_{\varepsilon,2}(\infty) \to \alpha_2(\infty) = 1$ as $\varepsilon \to 0$ also hold.

Ergodic theorems for singularly and super-singularly perturbed alternating processes have much more complex and interesting forms than for regularly perturbed alternating regenerative processes.

Functions $v_\varepsilon = p_{\varepsilon,12}^{-1} + p_{\varepsilon,21}^{-1}, \varepsilon \in (0, 1]$ and $w_\varepsilon = (p_{\varepsilon,12} + p_{\varepsilon,21})^{-1}, \varepsilon \in (0, 1]$ play important roles of so-called "time scaling" factors, respectively, for singularly and super-singularly perturbed models. In the case **(d)**, $0 < w_\varepsilon < v_\varepsilon < \infty$, for $\varepsilon \in (0, 1]$ and $w_\varepsilon, v_\varepsilon \to \infty$ as $\varepsilon \to 0$. In the cases **(e)** and **(f)**, $0 < w_\varepsilon < v_\varepsilon = \infty$, for $\varepsilon \in (0, 1]$ and $w_\varepsilon \to \infty$ as $\varepsilon \to 0$.

The main individual ergodic theorems for singularly perturbed alternating regenerative processes have forms of asymptotic relations, $P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,ij}^{(\beta)}(t, A)$ as $\varepsilon \to 0$, holding under assumption that condition $\mathbf{K}_\beta$ holds for some $\beta \in [0, \infty]$, for $A \in \Gamma, i, j = 1, 2$, and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that $t_\varepsilon/v_\varepsilon \to t \in [0, \infty]$ as $\varepsilon \to 0$.

The asymptotic behaviour of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ can differ for different asymptotic time zones determined by the asymptotic relation $t_\varepsilon/v_\varepsilon \to t \in [0, \infty]$. The corresponding limiting probabilities $\pi^{(\beta)}_{0,ij}(t, A)$ may depend on $t \in [0, \infty]$, parameter $\beta \in 0, \infty]$, appearing in condition $\mathbf{K}_\beta$, and, also, on the initial state $i \in \mathbb{Y}$, if $t \in [0, \infty)$. It is natural to classify the corresponding theorems as super-long, long, and short time ergodic theorem, respectively, for cases $t = \infty$, $t \in (0, \infty)$ and $t = 0$, for which the corresponding limiting probabilities take different analytical forms.

The corresponding individual ergodic theorems for super-singularly perturbed alternating regenerative processes have forms of analogous asymptotic relations, $P_{\varepsilon,ij}(t_\varepsilon, A) \to \dot{\pi}^{(\beta)}_{0,ij}(t, A)$ as $\varepsilon \to 0$, holding under assumption that condition $\mathbf{K}_0$ or $\mathbf{K}_\infty$ holds, for $A \in \Gamma$, $i, j = 1, 2$ and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that $t_\varepsilon/w_\varepsilon \to t \in [0, \infty]$ as $\varepsilon \to 0$.

In this case, the asymptotic behaviour of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ also can differ for different asymptotic time zones determined by the asymptotic relation $t_\varepsilon/w_\varepsilon \to t \in [0, \infty]$. The corresponding limiting probabilities $\dot{\pi}^{(\beta)}_{0,ij}(t, A)$ may depend on $t \in [0, \infty]$, parameter β, taking in this case one of two values 0 or ∞, and, also, on the initial state $i \in \mathbb{Y}$, if $t \in [0, \infty)$. As for singularly perturbed models, it is natural to classify the corresponding theorems as super-long, long and short time ergodic theorem, respectively, for cases $t = \infty$, $t \in (0, \infty)$ and $t = 0$, for which the corresponding limiting probabilities take different analytical forms.

Ergodic theorems for singularly perturbed models for the cases, where condition $\mathbf{K}_0$ or $\mathbf{K}_\infty$ is assumed to hold, can be compared with ergodic theorems for super-singularly perturbed models, respectively, for the cases **(e)** or **(f)**. Indeed, as was mentioned above, condition $\mathbf{K}_0$ or $\mathbf{K}_\infty$ holds, respectively, in the case **(e)** or **(f)**.

In cases **(e)** and **(f)**, i.e., for super-singularly perturbed models, $v_\varepsilon = \infty$, while $0 < w_\varepsilon < \infty$, for $\varepsilon \in (0, 1]$. The only factor w_ε can be used as a time scaling factor. In the case **(d)**, i.e., for singularly perturbed models, $0 < w_\varepsilon < v_\varepsilon < \infty$, for $\varepsilon \in (0, 1]$. The question arises if w_ε can be used as a time scaling factor instead of v_ε. The answer is in some sense affirmative, if condition $\mathbf{K}_\beta$ holds for some $\beta \in (0, \infty)$. Indeed, in this case, $w_\varepsilon/v_\varepsilon \to \beta(1 + \beta)^{-2} \in (0, \infty)$ as $\varepsilon \to 0$. The asymptotic relations, $t_\varepsilon/v_\varepsilon \to t$ as $\varepsilon \to 0$, and, $t_\varepsilon/w_\varepsilon \to t$ as $\varepsilon \to 0$, generate, in fact, in some sense equivalent asymptotic time zones. However, the answer for the above question is negative, if condition $\mathbf{K}_0$ or $\mathbf{K}_\infty$ holds. Indeed, in this case, $w_\varepsilon/v_\varepsilon \to 0$ as $\varepsilon \to 0$. The asymptotic relations, $t_\varepsilon/v_\varepsilon \to t$ as $\varepsilon \to 0$, and, $t_\varepsilon/w_\varepsilon \to t$ as $\varepsilon \to 0$, generate essentially different asymptotic time zones, in the corresponding ergodic theorems. This, actually, makes it possible to get, under the assumption that condition $\mathbf{K}_0$ or $\mathbf{K}_\infty$ holds, additional ergodic relations for singularly perturbed processes, similar to those given above for super-singularly perturbed processes, for asymptotic time zones generated by relation $t_\varepsilon/w_\varepsilon \to t$ as $\varepsilon \to 0$,

The extremal case, **(g)** $p_{\varepsilon,12}$, $p_{\varepsilon,21} = 0$, $\varepsilon \in [0, 1]$, corresponds to absolutely singular perturbed alternating regenerative processes. This case is not covered by condition $\mathbf{K}_\beta$. However, in this case the modulating process $\eta_\varepsilon(t) = \eta_\varepsilon(0)$, $t \ge 0$. Respectively, the process $\xi_\varepsilon(t)$, $t \ge 0$ coincides with the standard regenerative process $\xi_{\varepsilon,i}(t)$, $t \ge 0$, if $\eta_\varepsilon(0) = i$. The corresponding ergodic theorem for process $\xi_{\varepsilon,i}(t)$ is given by Theorem 3.1 for its particular case described in Sect. 3.2.3.

In conclusion, let us make some comments concerning ergodic theorems for probabilities $P_{\varepsilon,\bar{p}_\varepsilon,j}(t, A) = \mathsf{P}\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = j\} = p_{\varepsilon,1}P_{\varepsilon,1j}(t, A) + p_{\varepsilon,2}P_{\varepsilon,2j}(t, A)$.

In models, where the corresponding limits for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ do not depend of the initial state i, for example, for regularly perturbed alternating regenerative processes, probabilities $P_{\varepsilon,\bar{p}_\varepsilon,j}(t_\varepsilon, A)$ converge to the same limits for any initial distributions $\bar{p}_\varepsilon = \langle p_{\varepsilon,1}, p_{\varepsilon,2}\rangle$.

However, in models, where the corresponding limits for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ may depend of the initial state i, for example, for some singularly or super-singularly perturbed alternating regenerative processes, probabilities $P_{\varepsilon,\bar{p}_\varepsilon,j}(t, A)$ converge to some limits under an additional condition of asymptotic stability for initial distributions:

L: $p_{\varepsilon,i} \to p_{0,i}$ as $\varepsilon \to 0$, for $i = 1, 2$.

If, for example, condition **L** holds and, $P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi^{(\beta)}_{0,ij}(t, A)$ as $\varepsilon \to 0$, for $i = 1, 2$, then,

$$P_{\varepsilon,\bar{p}_\varepsilon,j}(t_\varepsilon, A) \to \pi^{(\beta)}_{0,\bar{p}_0,j}(t, A) = p_{0,1}\pi^{(\beta)}_{0,1j}(t, A) + p_{0,2}\pi^{(\beta)}_{0,2j}(t, A) \text{ as } \varepsilon \to 0. \tag{3.21}$$

3.2.7 Aggregation of Regeneration Times

The alternating regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t)), t \geq 0$ is a standard regenerative process with regeneration times $\tau_{\varepsilon,0}, \tau_{\varepsilon,1}, \tau_{\varepsilon,2}, \ldots$ if and only if the joint distributions of random variables $\xi_{\varepsilon,i,n}(t_k), k = 1, \ldots, r$ and $\kappa_{\varepsilon,i,n}, \eta_{\varepsilon,i,n}$ do not depend on $n \geq 1$, for every $t_k \in [0, \infty), k = 1, \ldots, r, r \geq, i = 1, 2$.

However, it is possible to construct new aggregated regeneration times such that the process $(\xi_\varepsilon(t), \eta_\varepsilon(t)), t \geq 0$ becomes a standard regenerative process with these new regeneration times.

Let us define stopping times for Markov chain $\eta_{\varepsilon,n}$ that are, $\hat{\theta}_\varepsilon[r] = \min(k > r : \eta_{\varepsilon,k} = \eta_{\varepsilon,r})$, which is the first after r return time to the state $\eta_{\varepsilon,r}$, $\tilde{\theta}_\varepsilon[r] = \min(k > r : \eta_{\varepsilon,k} \neq \eta_{\varepsilon,r})$, which is the first after r time of change of state $\eta_{\varepsilon,r}$, and $\check{\theta}_\varepsilon[r] = \min(k > \tilde{\theta}_\varepsilon[r] : \eta_{\varepsilon,k} = \eta_{\varepsilon,r})$, which is the first after $\tilde{\theta}_\varepsilon[r]$ return time to the state $\eta_{\varepsilon,r}$. Obviously, the above return times are connected by the inequality $r < \hat{\theta}_\varepsilon[r] < \check{\theta}_\varepsilon[r]$, for $r = 0, 1, \ldots$.

Let us also $\hat{\nu}_{\varepsilon,0} = 0$, $\hat{\nu}_{\varepsilon,n} = \hat{\theta}[\hat{\nu}_{\varepsilon,n-1}], n = 1, 2, \ldots$, and $\check{\nu}_{\varepsilon,0} = 0$, $\check{\nu}_{\varepsilon,n} = \check{\theta}[\check{\nu}_{\varepsilon,n-1}]$, $n = 1, 2, \ldots$ be the corresponding sequential return times to the state $\eta_{\varepsilon,0}$ by the Markov chain $\eta_{\varepsilon,n}$.

Let also us consider sequential return times $\hat{\tau}_{\varepsilon,n} = \tau_{\varepsilon,\hat{\nu}_{\varepsilon,n}}, n = 0, 1, \ldots$ and $\check{\tau}_{\varepsilon,n} = \tau_{\varepsilon,\check{\nu}_{\varepsilon,n}}, n = 0, 1, \ldots$ to the state $\eta_\varepsilon(0)$ by the semi-Markov process $\eta_\varepsilon(t)$.

Process $(\xi_\varepsilon(t), \eta_\varepsilon(t)), t \geq 0$ is a regenerative process with regeneration times $\hat{\tau}_{\varepsilon,n}, n = 0, 1, \ldots$. It also is a regenerative process with regeneration times $\check{\tau}_{\varepsilon,n}, n = 0, 1, \ldots$.

We can also consider shifted sequences of discrete time stopping times $\hat{\nu}'_{\varepsilon,0} = 0$, $\hat{\nu}'_{\varepsilon,1} = \tilde{\theta}_\varepsilon[0]$, $\hat{\nu}'_{\varepsilon,n} = \hat{\theta}[\hat{\nu}'_{\varepsilon,n-1}]$, $n = 2, 3, \ldots$ and $\check{\nu}'_{\varepsilon,0} = 0$, $\check{\nu}'_{\varepsilon,1} = \tilde{\theta}_\varepsilon[0]$, $\check{\nu}'_{\varepsilon,n} = \check{\theta}[\check{\nu}'_{\varepsilon,n-1}]$, $n = 2, 3, \ldots$, and the corresponding stopping times $\hat{\tau}'_{\varepsilon,n} = \tau_{\varepsilon,\hat{\nu}'_{\varepsilon,n}}$, $n = 0, 1, \ldots$ and $\check{\tau}'_{\varepsilon,n} = \tau_{\varepsilon,\check{\nu}'_{\varepsilon,n}}$, $n = 0, 1, \ldots$.

If $\eta_\varepsilon(0) = 1$, then the stopping times $\hat{\tau}_{\varepsilon,n}$, $n = 1, 2, \ldots$ and $\check{\tau}_{\varepsilon,n}$, $n = 1, 2, \ldots$ are return times to the state 1 for the semi-Markov process $\eta_\varepsilon(t)$. As far as the shifted stopping times $\hat{\tau}'_{\varepsilon,n}$ and $\check{\tau}'_{\varepsilon,n}$ are concerned, $\hat{\tau}'_{\varepsilon,1} = \check{\tau}'_{\varepsilon,1}$ is the first hitting time to state 2, while $\hat{\tau}_{\varepsilon,n}$, $n = 2, 3, \ldots$ and $\check{\tau}_{\varepsilon,n}$, $n = 2, 3, \ldots$ are return times to the state 2 for the semi-Markov process $\eta_\varepsilon(t)$.

If $\eta_\varepsilon(0) = 2$, then the stopping times $\hat{\tau}_{\varepsilon,n}$, $n = 1, 2, \ldots$ and $\check{\tau}_{\varepsilon,n}$, $n = 1, 2, \ldots$ are return times to the state 2 for the semi-Markov process $\eta_\varepsilon(t)$. As far as the shifted stopping times $\hat{\tau}'_{\varepsilon,n}$ and $\check{\tau}'_{\varepsilon,n}$ are concerned, $\hat{\tau}'_{\varepsilon,1} = \check{\tau}'_{\varepsilon,1}$ is the first hitting time to state 1, while $\hat{\tau}_{\varepsilon,n}$, $n = 2, 3, \ldots$ and $\check{\tau}_{\varepsilon,n}$, $n = 2, 3, \ldots$ are return times to the state 1 for the semi-Markov process $\eta_\varepsilon(t)$.

Process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \geq 0$ is a regenerative process with the transition period $[0, \hat{\tau}'_{\varepsilon,1})$ and the regeneration times $\hat{\tau}'_{\varepsilon,n}$, $n = 0, 1, \ldots$. It also is a regenerative process with the transition period $[0, \check{\tau}_{\varepsilon,1})$ and the regeneration times $\check{\tau}'_{\varepsilon,n}$, $n = 0, 1, \ldots$.

We shall see that aggregated regeneration times $\hat{\tau}_{\varepsilon,n}$ and $\hat{\tau}'_{\varepsilon,n}$ work well in ergodic theorems for models with regular perturbations. However, these regeneration times do not work well for the models with singular and super-singular perturbations. Here, the regeneration times $\check{\tau}_{\varepsilon,n}$ and $\check{\tau}'_{\varepsilon,n}$ should be used.

3.3 Ergodic Theorems for Regularly Perturbed Alternating Regenerative Processes

In this section, we present individual ergodic theorems for regularly perturbed alternating regenerative processes. These theorems are, in fact, rather simple examples illustrating applications of results generalising the renewal theorem to the model of perturbed renewal equation given in [43–45] and individual ergodic theorems for perturbed regenerative processes throughly presented in [14]. Other related references are given in the introduction.

3.3.1 Perturbed Standard Alternating Regenerative Processes

Let us consider regularly perturbed standard alternating regenerative processes, where, additionally to **F**–**J**, the following condition holds:

M$_1$: $p_{\varepsilon,12}$, $p_{\varepsilon,21} = 1$, for $\varepsilon \in [0, 1]$.

In this case, the Markov chain $\eta_{0,n}$ is ergodic. Obviously, parameter $\beta = 1$, and its stationary probabilities are, $\alpha_1(1) = \alpha_2(1) = \frac{1}{2}$.

Conditions **F–I** and $\mathbf{M}_1$ imply that the semi-Markov process $\eta_0(t)$ is ergodic. Its stationary probabilities have the form,

$$\rho_1(1) = e_{0,1}/(e_{0,1} + e_{0,2}),\ \rho_2(1) = e_{0,2}/(e_{0,1} + e_{0,2}). \tag{3.22}$$

The corresponding stationary probabilities for the alternating regenerative process $(\xi_0(t), \eta_0(t))$ have the form,

$$\pi^{(1)}_{0,j}(A) = \rho_j(1)\pi_{0,j}(A),\ A \in \mathcal{B}_{\mathbb{X}},\ j = 1, 2. \tag{3.23}$$

The ergodic theorem for perturbed standard alternating regenerative processes takes the following form.

Theorem 3.4 *Let conditions* **F–J** *and* $\mathbf{M}_1$ *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and any* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi^{(1)}_{0,j}(A) \text{ as } \varepsilon \to 0. \tag{3.24}$$

Proof In the case, where condition $\mathbf{M}_1$ holds, the stopping times $\tilde{\theta}[r] = r + 1$ and $\hat{\theta}[r] = \check{\theta}[r] = r + 2$, for $r = 0, 1, \ldots$.

Thus, the regenerations times $\hat{\tau}_{\varepsilon,n} = \check{\tau}_{\varepsilon,n} = \tau_{\varepsilon,2n}$, $n = 0, 1, \ldots$ and

$$\hat{\tau}'_{\varepsilon,0} = \check{\tau}'_{\varepsilon,0} = 0,\ \hat{\tau}'_{\varepsilon,1} = \check{\tau}'_{\varepsilon,1} = \tau_{\varepsilon,1},\ \hat{\tau}'_{\varepsilon,n} = \check{\tau}'_{\varepsilon,n} = \tau_{\varepsilon,2n-1},\ n = 2, 3, \ldots.$$

Here and henceforth, we use the same symbol for equalities or inequalities which hold for random variables, for all $\omega \in \Omega_\varepsilon$ or almost sure, since this difference does not affect the corresponding probabilities and expectations.

Therefore, the standard alternating regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \ge 0$ is a standard regenerative process with regeneration times $\tau_{\varepsilon,0}$, $\tau_{\varepsilon,2}$, $\tau_{\varepsilon,4}$, ... It also can be considered as a regenerative process with transition period $[0, \tau_{\varepsilon,1})$ and regenerative times $\tau_{\varepsilon,0}$, $\tau_{\varepsilon,1}$, $\tau_{\varepsilon,3}$, $\tau_{\varepsilon,5}$, ...

Regenerative lifetimes are not involved. We can use the Theorems 3.1–3.3, for the model with stopping probabilities $f_\varepsilon = 0$, $\varepsilon \in [0, 1]$.

First, let us analyse the asymptotic behaviour of probabilities $P_{\varepsilon,11}(t, A)$. In this case, we do prefer to consider $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \ge 0$ as the standard regenerative process with regeneration times $\tau_{\varepsilon,0}$, $\tau_{\varepsilon,2}$, ...

The renewal type Eq. (3.7) takes for probabilities $P_{\varepsilon,11}(t, A)$ the following form,

$$P_{\varepsilon,11}(t, A) = q^{(2)}_{\varepsilon,1}(t, A) + \int_0^t P_{\varepsilon,11}(t - s, A)Q^{(2)}_{\varepsilon,11}(ds),\ t \ge 0, \tag{3.25}$$

where $q^{(2)}_{\varepsilon,1}(t, A) = \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \tau_{\varepsilon,2} > t\}$, $t \ge 0$ and

$$Q^{(2)}_{\varepsilon,11}(t) = \mathsf{P}_1\{\tau_{\varepsilon,2} \le t\},\ t \ge 0.$$

In this case, $\eta_\varepsilon(t) = 1$, for $t \in [0, \tau_{\varepsilon,1})$, and $\eta_\varepsilon(t) = 2$, for $t \in [\tau_{\varepsilon,1}, \tau_{\varepsilon,2})$. Therefore, for every $A \in \mathcal{B}_{\mathbb{X}}$, $t \geq 0$,

$$\begin{aligned} q^{(2)}_{\varepsilon,1}(t, A) &= \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \tau_{\varepsilon,2} > t\} \\ &= \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \tau_{\varepsilon,1} > t\} = q_{\varepsilon,1}(t, A), \end{aligned} \tag{3.26}$$

Also, for $t \geq 0$,

$$Q^{(2)}_{\varepsilon,11}(t) = \mathsf{P}_1\{\tau_{\varepsilon,2} \leq t\} = Q_{\varepsilon,12}(t) * Q_{\varepsilon,21}(t), \tag{3.27}$$

and, thus,

$$e^{(2)}_{\varepsilon,11} = \mathsf{E}_1\tau_{\varepsilon,2} = e_{\varepsilon,12} + e_{\varepsilon,21}. \tag{3.28}$$

Note that condition $\mathbf{M}_1$ implies that expectations $e_{\varepsilon,11}, e_{\varepsilon,22} = 0$ and, therefore, $e_{\varepsilon,12} + e_{\varepsilon,21} = e_{\varepsilon,1} + e_{\varepsilon,2}$.

Condition **F** obviously implies that condition **A** holds. Relation (3.27) and conditions **G**, **H**, and $\mathbf{M}_1$ imply that condition **B (a)** holds. Relation (3.27) and condition **H (b)** implies that condition **B (b)** holds. Relation (3.28) and condition **I** imply that condition **C** holds. Relation (3.26) and condition **J** imply that condition **D** holds. As was mentioned above, in this case, $f_\varepsilon \equiv 0$. Thus, all conditions of Theorem 3.1 holds, and the ergodic relation given in this theorem takes place for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$. In this case, it takes the form of relation (3.24), where one should choose $i, j = 1$.

Second, let us analyse the asymptotic behaviour of probabilities $P_{\varepsilon,21}(t, A)$. In this case, we do prefer to consider $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \geq 0$ as the regenerative process with transition period $[0, \tau_{\varepsilon,1})$ and regenerative times $\tau_{\varepsilon,0}, \tau_{\varepsilon,1}, \tau_{\varepsilon,3}, \ldots$.

The shifted process $(\xi_\varepsilon(\tau_{\varepsilon,1} + t), \eta_\varepsilon(\tau_{\varepsilon,1} + t))$, $t \geq 0$ is a standard regenerative process. If $\eta_\varepsilon(0) = 2$, then $\eta_\varepsilon(\tau_{\varepsilon,1}) = 1$. That is why, probabilities $P_{\varepsilon,11}(t, A)$ play for the above shifted regenerative process the role of probabilities $P^{(1)}_\varepsilon(t, A)$ defined in Sect. 3.2.1.

The distribution function for the duration of the transition period $[0, \tau_{\varepsilon,1})$ has, in this case, the following form,

$$\mathsf{P}_2\{\tau_{\varepsilon,1} \leq t\} = Q_{\varepsilon,21}(t), \ t \geq 0. \tag{3.29}$$

Relation (3.29) and conditions **H**, $\mathbf{M}_1$ imply that condition **E** holds. Thus, all conditions of Theorem 3.2 hold, and the corresponding ergodic relation for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ also holds for probabilities $P_{\varepsilon,21}(t_\varepsilon, A)$.

Due to the symmetricity of conditions **F**–**J** and $\mathbf{M}_1$ with respect to the indices $i, j = 1, 2$, the ergodic relations, analogous to the mentioned above ergodic relations for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ and $P_{\varepsilon,21}(t_\varepsilon, A)$, also take place for probabilities $P_{\varepsilon,22}(t_\varepsilon, A)$ and $P_{\varepsilon,12}(t_\varepsilon, A)$. The only, stationary probabilities $\pi^{(1)}_{0,1}(A)$ should be replaced by stationary probabilities $\pi^{(1)}_{0,2}(A)$ in the corresponding ergodic relations. □

3.3.2 Regularly Perturbed Alternating Regenerative Processes

Let us now consider alternating regenerative processes with a regular perturbation model, where additionally to **F–J**, the following condition holds:

$\mathbf{M}_2$: $p_{0,12},\ p_{0,21} > 0$.

Note that condition $\mathbf{M}_1$ is a particular case of condition $\mathbf{M}_2$, and, thus, any standard alternating regenerative process also is a regularly alternating regenerative process.

In this case, the Markov chain $\eta_{0,n}$ is ergodic. Parameter $\beta = p_{0,12}/p_{0,21} \in (0, \infty)$, and the stationary probabilities for the above Markov chain are, $\alpha_1(\beta) = \frac{1}{1+\beta}$ and $\alpha_2(\beta) = \frac{1}{1+\beta^{-1}}$.

Conditions **F–I** and $\mathbf{M}_2$ imply that the semi-Markov process $\eta_0(t)$ is ergodic. Its stationary probabilities have the form,

$$\rho_1(\beta) = \frac{e_{0,1}\alpha_1(\beta)}{e_{0,1}\alpha_1(\beta) + e_{0,2}\alpha_2(\beta)},\ \rho_2(\beta) = \frac{e_{0,2}\alpha_2(\beta)}{e_{0,1}\alpha_1(\beta) + e_{0,2}\alpha_2(\beta)}. \tag{3.30}$$

The corresponding stationary probabilities for the alternating regenerative process $\xi_0(t)$ has the form, for $\beta \in (0, \infty)$,

$$\pi_{0,j}^{(\beta)}(A) = \rho_j(\beta)\pi_{0,j}(A),\ A \in \mathcal{B}_{\mathbb{X}},\ j = 1, 2. \tag{3.31}$$

The ergodic theorem for perturbed alternating regenerative processes takes the following form.

Theorem 3.5 *Let conditions* **F–J** *hold and, also, condition* $\mathbf{M}_2$ *holds and parameter* $p_{0,12}/p_{0,21} = \beta \in (0, \infty)$. *Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and any* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,j}^{(\beta)}(A) \text{ as } \varepsilon \to 0. \tag{3.32}$$

Proof As was pointed out in Sect. 3.3.1, process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$ is a regenerative process with regeneration times with regeneration times $\hat{\tau}_{\varepsilon,n}$, $n = 0, 1, \ldots$. It is also a regenerative process with the transition period $[0, \hat{\tau}'_{\varepsilon,1})$ and the regeneration times $\hat{\tau}'_{\varepsilon,n}$, $n = 0, 1, \ldots$.

Again, regenerative lifetimes are not involved. We can use the Theorems 3.1–3.3, for the model with stopping probabilities $f_\varepsilon = 0$, $\varepsilon \in [0, 1]$.

First, let us analyse the asymptotic behaviour of probabilities $P_{\varepsilon,11}(t, A)$. In this case, we do prefer to consider $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \ge 0$ as the standard regenerative process with regeneration times $\hat{\tau}_{\varepsilon,n}$, $n = 0, 1, \ldots$.

The renewal type Eq. (3.7) for probabilities $P_{\varepsilon,11}(t, A)$ takes, in this case, the following form,

$$P_{\varepsilon,11}(t, A) = \hat{q}_{\varepsilon,1}(t, A) + \int_0^t P_{\varepsilon,11}(t - s, A)\hat{Q}_{\varepsilon,11}(ds), t \ge 0, \tag{3.33}$$

where $\hat{q}_{\varepsilon,1}(t, A) = \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \hat{\tau}_{\varepsilon,1} > t\}, t \geq 0$ and $\hat{Q}_{\varepsilon,11}(t) = \mathsf{P}_1\{\hat{\tau}_{\varepsilon,1} \leq t\}, t \geq 0$.

If $\eta_\varepsilon(0) = 1$, then $\eta_\varepsilon(t) = 1$ for $t \in [0, \tau_{\varepsilon,1})$. Also, $\hat{\tau}_{\varepsilon,1} = \tau_{\varepsilon,1}$, if $\eta_{\varepsilon,1} = 1$, and $\eta_\varepsilon(t) = 2$, for $t \in [\tau_{\varepsilon,1}, \hat{\tau}_{\varepsilon,1})$, if $\eta_{\varepsilon,1} = 2$. Therefore, for every $A \in \mathcal{B}_{\mathbb{X}}, t \geq 0$,

$$\begin{aligned}\hat{q}_{\varepsilon,1}(t, A) &= \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \hat{\tau}_{\varepsilon,1} > t\} \\ &= \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \tau_{\varepsilon,1} > t, \eta_{\varepsilon,1} = 1\} \\ &\quad + \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \tau_{\varepsilon,1} > t, \eta_{\varepsilon,1} = 2\} \\ &= \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \tau_{\varepsilon,1} > t\} = q_{\varepsilon,1}(t, A). \end{aligned} \tag{3.34}$$

In this case, $\hat{Q}_{\varepsilon,11}(t)$ is the distribution function of the first return time to state 1 for semi-Markov process $\eta_\varepsilon(t)$. It can be expressed in terms of convolutions of transition probabilities for this semi-Markov process,

$$\hat{Q}_{\varepsilon,11}(t) = Q_{\varepsilon,11}(t) + Q_{\varepsilon,12}(t) * \sum_{n=0}^{\infty} Q_{\varepsilon,22}^{*n}(t) * Q_{\varepsilon,21}(t), t \geq 0. \tag{3.35}$$

Relation (3.35) takes the following equivalent form in terms of Laplace transforms,

$$\begin{aligned}\hat{\phi}_{\varepsilon,11}(s) &= \int_0^\infty e^{-st} \hat{Q}_{\varepsilon,11}(dt) \\ &= \phi_{\varepsilon,11}(s) + \phi_{\varepsilon,12}(s) \sum_{n=0}^{\infty} \phi_{\varepsilon,22}^n(s) \phi_{\varepsilon,21}(s) \\ &= \phi_{\varepsilon,11}(s) + \phi_{\varepsilon,12}(s) \frac{1}{1 - \phi_{\varepsilon,22}(s)} \phi_{\varepsilon,21}(s), \ s \geq 0. \end{aligned} \tag{3.36}$$

Relation (3.35) also implies that random variable $\hat{\nu}_{\varepsilon,1}$ has a so-called burned geometric distribution that is,

$$\hat{\nu}_{\varepsilon,1} = \begin{cases} 1 & \text{with probability } p_{\varepsilon,11}, \\ n & \text{with probability } p_{\varepsilon,12}\, p_{\varepsilon,22}^{n-2}\, p_{\varepsilon,21}, \text{ for } n \geq 2. \end{cases}$$

This fact and conditions **G**, **H**, and $\mathbf{M}_2$ imply, in an obvious way, that expectation $\hat{e}_{\varepsilon,11} = \mathsf{E}_1 \hat{\tau}_{\varepsilon,1} < \infty$. It can be easily computed, for example, using the derivative of the Laplace transform $\hat{\phi}_{\varepsilon,11}(s)$ at zero,

$$\begin{aligned}\hat{e}_{\varepsilon,11} = \mathsf{E}_1 \hat{\tau}_{\varepsilon,1} &= -\hat{\phi}'_{\varepsilon,11}(0) = e_{\varepsilon,11} + e_{\varepsilon,12} \frac{1}{1 - p_{\varepsilon,22}} p_{\varepsilon,21} \\ &\quad + p_{\varepsilon,12} \frac{e_{\varepsilon,22}}{(1 - p_{\varepsilon,22})^2} p_{\varepsilon,21} + p_{\varepsilon,12} \frac{1}{1 - p_{\varepsilon,22}} e_{\varepsilon,21} \\ &= \frac{e_{\varepsilon,1} p_{\varepsilon,21} + e_{\varepsilon,2} p_{\varepsilon,12}}{p_{\varepsilon,21}} \\ &= \frac{e_{\varepsilon,1} \alpha_1(\beta_\varepsilon) + e_{\varepsilon,2} \alpha_2(\beta_\varepsilon)}{\alpha_1(\beta_\varepsilon)}. \end{aligned} \tag{3.37}$$

Obviously, $\hat{\tau}_{\varepsilon,n} \geq \tau_{\varepsilon,n}$, for $n = 0, 1, \ldots$. Thus, condition **F** implies that condition **A** holds. Relations (3.35), (3.36) and conditions **G**, **H**, and $\mathbf{M}_2$ imply that Laplace transforms $\hat{\phi}_{\varepsilon,11}(s) \to \hat{\phi}_{0,11}(s)$ as $\varepsilon \to 0$, for $s \geq 0$. Thus, by Remark 3.1, condition **B (a)** holds. Also, relation (3.35) and condition **H (b)** implies that condition **B (b)** holds. Relation (3.37) and conditions **H** and **I** imply that condition **C** holds. Relation (3.34) and condition **J** imply that condition **D** holds. As was mentioned above, in this case, $f_\varepsilon \equiv 0$. Thus, all conditions of Theorem 3.1 hold, and the ergodic relation given in this theorem takes place for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$. In this case, it takes the form of relation (3.32), where one should choose $i, j = 1$, i.e., for every $A \in \Gamma$, and any $0 \leq t_\varepsilon \to \infty$ as $\varepsilon \to 0$,

$$\begin{aligned} P_{\varepsilon,11}(t_\varepsilon, A) &\to \frac{\alpha_1(\beta)}{e_{0,1}\alpha_1(\beta) + e_{0,2}\alpha_2(\beta)} \int_0^\infty q_{0,1}(s, A) m(ds) \\ &= \frac{e_{0,1}\alpha_1(\beta)}{e_{0,1}\alpha_1(\beta) + e_{0,2}\alpha_2(\beta)} \frac{1}{e_{0,1}} \int_0^\infty q_{0,1}(s, A) m(ds) \\ &= \rho_j(\beta)\pi_{0,j}(A) = \pi_{0,j}^{(\beta)}(A) \text{ as } \varepsilon \to 0. \end{aligned} \tag{3.38}$$

Second, let us analyse the asymptotic behaviour of probabilities $P_{\varepsilon,21}(t, A)$. In this case, we do prefer to consider $(\xi_\varepsilon(t), \eta_\varepsilon(t)), t \geq 0$ as the regenerative process with transition period $[0, \hat{\tau}'_{\varepsilon,1})$ and regenerative times $\hat{\tau}'_{\varepsilon,0}, \hat{\tau}'_{\varepsilon,1} = \tilde{\tau}_{\varepsilon,1}, \hat{\tau}'_{\varepsilon,2}, \hat{\tau}'_{\varepsilon,3}, \ldots$.

The shifted process $(\xi_\varepsilon(\hat{\tau}'_{\varepsilon,1} + t), \eta_\varepsilon(\hat{\tau}'_{\varepsilon,1} + t)), t \geq 0$ is a standard regenerative process. If $\eta_\varepsilon(0) = 2$, then $\eta_\varepsilon(\tilde{\tau}_{\varepsilon,1}) = 1$. That is why, probabilities $P_{\varepsilon,11}(t, A)$ play for this process the role of probabilities $P_\varepsilon^{(1)}(t, A)$ pointed out in Sect. 3.2.1.

The distribution function for the duration of the transition period $[0, \tilde{\tau}_{\varepsilon,1})$ has, in this case, the following form,

$$\begin{aligned} \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \leq t\} &= \tilde{Q}_{\varepsilon,21}(t) \\ &= \sum_{n=0}^\infty Q_{\varepsilon,22}^{*n}(t) * Q_{\varepsilon,21}(t), t \geq 0. \end{aligned} \tag{3.39}$$

This relation takes the following equivalent form in terms of Laplace transforms,

$$\begin{aligned} \tilde{\phi}_{\varepsilon,21}(s) &= \int_0^\infty e^{-st} \tilde{Q}_{\varepsilon,21}(dt) \\ &= \sum_{n=0}^\infty \phi_{\varepsilon,22}^n(s)\phi_{\varepsilon,21}(s) \\ &= \frac{\phi_{\varepsilon,21}(s)}{1 - \phi_{\varepsilon,22}(s)}, \ s \geq 0. \end{aligned} \tag{3.40}$$

Relations (3.39), (3.40) and conditions **G**, **H**, and $\mathbf{M}_2$ imply that Laplace transforms $\tilde{\phi}_{\varepsilon,21}(s) \to \tilde{\phi}_{0,21}(s)$ as $\varepsilon \to 0$, for $s \geq 0$. Thus, by Remark 3.1, condition **E** holds. All conditions of Theorem 3.2 hold, and the corresponding ergodic relation for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ also holds for probabilities $P_{\varepsilon,21}(t_\varepsilon, A)$.

Due to the symmetricity of conditions **F–J** and $\mathbf{M}_2$ with respect to the indices $i, j = 1, 2$, the ergodic relations, analogous to the mentioned above ergodic relations for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ and $P_{\varepsilon,21}(t_\varepsilon, A)$, also take place for probabilities $P_{\varepsilon,22}(t_\varepsilon, A)$ and $P_{\varepsilon,12}(t_\varepsilon, A)$. The only, the stationary probabilities $\pi_{0,1}^{(\beta)}(A)$ should be replaced by stationary probabilities $\pi_{0,2}^{(\beta)}(A)$ in the corresponding ergodic relations. □

Remark 3.2 Theorem 3.4 is a particular case of Theorem 3.5. In this case, the ergodic relation (3.32) takes the form of ergodic relation (3.24).

3.3.3 *Semi-regularly Perturbed Alternating Regenerative Processes*

Let us now consider alternating regenerative processes with the semi-regular perturbation model, where additionally to **F–J**, the following condition holds:

$\mathbf{M}_3$: **(a)** $p_{0,12} = 0$, $p_{0,21} > 0$ or **(b)** $p_{0,12} > 0$, $p_{0,21} = 0$.

In this case, the Markov chain $\eta_{0,n}$ is ergodic. Parameter $\beta = p_{0,12}/p_{0,21} = 0$, and the stationary probabilities for the above Markov chain are, $\alpha_1(0) = 1$, $\alpha_2(0) = 0$, if condition $\mathbf{M}_3$ **(a)** holds. While, $\beta = p_{0,12}/p_{0,21} = \infty$, and the stationary probabilities for the above Markov chain are, $\alpha_1(\infty) = 0$, $\alpha_2(\infty) = 1$, if condition $\mathbf{M}_3$ **(b)** holds.

Conditions **F–J** and $\mathbf{M}_3$ imply that the semi-Markov process $\eta_0(t)$ is ergodic. Its stationary probabilities have the form,

$$\rho_1(0) = e_{0,1}\alpha_1(0)/(e_{0,1}\alpha_1(0) + e_{0,2}\alpha_2(0)) = 1,$$
$$\rho_2(0) = \varepsilon_{0,2}\alpha_2(0)/(e_{0,1}\alpha_1(0) + e_{0,2}\alpha_2(0)) = 0,$$

if condition $\mathbf{M}_3$ **(a)** holds. While, its stationary probabilities have the form,

$$\rho_1(\infty) = e_{0,1}\alpha_1(\infty)/(e_{0,1}\alpha_1(\infty) + e_{0,2}\alpha_2(\infty)) = 0,$$
$$\rho_2(\infty) = e_{0,2}\alpha_2(\infty)/(e_{0,1}\alpha_1(\infty) + e_{0,2}\alpha_2(\infty)) = 1,$$

if condition $\mathbf{M}_3$ **(b)** holds. The corresponding stationary probabilities for the alternating regenerative process $(\xi_0(t), \eta_0(t))$ have the form,

$$\pi_{0,j}^{(\beta)}(A) = \rho_j(\beta)\pi_{0,j}(A),\ A \in \mathcal{B}_{\mathbb{X}},\ j = 1, 2,\ \text{for } \beta = 0 \text{ and } \beta = \infty,\ \text{, i.e.,}$$

$$\pi_{0,j}^{(0)}(A) = \begin{cases} \pi_{0,1}(A) & \text{for } j = 1, \\ 0 & \text{for } j = 2, \end{cases} \tag{3.41}$$

and

$$\pi_{0,j}^{(\infty)}(A) = \begin{cases} 0 & \text{for } j = 1, \\ \pi_{0,2}(A) & \text{for } j = 2. \end{cases} \tag{3.42}$$

The ergodic theorems for perturbed alternating regenerative processes take the following forms.

Theorem 3.6 *Let conditions* **F–J** *and* $\mathbf{M}_3$ **(a)** *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and any* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,j}^{(0)}(A) \text{ as } \varepsilon \to 0. \tag{3.43}$$

Theorem 3.7 *Let conditions* **F–J** *and* $\mathbf{M}_3$ **(b)** *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and any* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,j}^{(\infty)}(A) \text{ as } \varepsilon \to 0. \tag{3.44}$$

Proof Process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$ is a standard regenerative process with regeneration times $\hat{\tau}_{\varepsilon,n}, n = 0, 1, \ldots$. It also is a regenerative process with transition period $[0, \hat{\tau}'_{\varepsilon,1})$ and regenerative times $\hat{\tau}'_{\varepsilon,n}, n = 0, 1, \ldots$.

Again, regenerative stopping is not involved. We can use the Theorems 3.1–3.3, for the model with stopping probabilities $f_\varepsilon = 0, \varepsilon \in [0, 1]$.

Let us consider the case, where condition $\mathbf{M}_3$ **(a)** holds.

Let us analyse the asymptotic behaviour for probabilities $P_{\varepsilon,1j}(t, A)$, $j = 1, 2$. In this case, we prefer to consider $(\xi_\varepsilon(t), \eta_\varepsilon(t)), t \ge 0$ as the standard regenerative process with regeneration times $\hat{\tau}_{\varepsilon,0}, \hat{\tau}_{\varepsilon,1}, \hat{\tau}_{\varepsilon,2}, \ldots$.

First, let us analyse the asymptotic behaviour of probabilities $P_{\varepsilon,11}(t, A)$.

The renewal type Eq. (3.7) takes for probabilities $P_{\varepsilon,1j}(t, A)$ the following form, for $j = 1, 2$,

$$P_{\varepsilon,1j}(t, A) = \hat{q}_{\varepsilon,1j}(t, A) + \int_0^t P_{\varepsilon,1j}(t - s, A)\hat{Q}_{\varepsilon,11}(ds), t \ge 0, \tag{3.45}$$

where $\hat{q}_{\varepsilon,1j}(t, A) = \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = j, \hat{\tau}_{\varepsilon,1} > t\}, t \ge 0, j = 1, 2$ and $\hat{Q}_{\varepsilon,11}(t) = \mathsf{P}_1\{\hat{\tau}_{\varepsilon,1} \le t\}, t \ge 0$.

In the case of probabilities $P_{\varepsilon,11}(t, A)$, we can repeat all calculations made in relations (3.33)–(3.37), given in the proof of Theorem 3.5. These relations, in fact, take simpler forms.

Analogously to relation (3.34), one can get, for every $A \in \mathcal{B}_{\mathbb{X}}, t \ge 0$,

$$\hat{q}_{\varepsilon,11}(t, A) = \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \tau_{\varepsilon,1} > t\} = q_{\varepsilon,1}(t, A). \tag{3.46}$$

Also, as was pointed out in comments related to relation (3.35), $\hat{Q}_{\varepsilon,11}(t)$ is the distribution function of the first return time to state 1 for semi-Markov process $\eta_\varepsilon(t)$, and the following formula, analogous to (3.36), takes place for its Laplace transform,

$$\hat{\phi}_{\varepsilon,11}(s) = \int_0^\infty e^{-st}\,\hat{Q}_{\varepsilon,11}(dt)$$
$$= \phi_{\varepsilon,11}(s) + \phi_{\varepsilon,12}(s)\frac{1}{1-\phi_{\varepsilon,22}(s)}\phi_{\varepsilon,21}(s),\ s \geq 0. \tag{3.47}$$

Also, the following formula, analogous to (3.37), takes place for expectations,

$$\hat{e}_{\varepsilon,11} = \mathsf{E}_1\hat{\tau}_{\varepsilon,1} = -\hat{\phi}'_{\varepsilon,11}(0) = \frac{e_{\varepsilon,1}\alpha_1(\beta_\varepsilon) + e_{\varepsilon,2}\alpha_2(\beta_\varepsilon)}{\alpha_1(\beta_\varepsilon)}. \tag{3.48}$$

Conditions **G**, **H**, and $\mathbf{M}_3$ **(a)** imply that, in relation (3.47), either Laplace transform $\phi_{\varepsilon,12}(s) = 0$, for $s \geq 0$, if $p_{\varepsilon,12} = 0$, for $\varepsilon \in [0, 1]$, or $\phi_{\varepsilon,12}(s) \to 0$ as $\varepsilon \to 0$, for $s \geq 0$, if $0 < p_{\varepsilon,12} \to 0$ as $\varepsilon \to 0$. This implies that the Laplace transforms $\hat{\phi}_{\varepsilon,11}(s) \to \hat{\phi}_{0,11}(s) = \phi_{0,11}(s)$ as $\varepsilon \to 0$, for $s \geq 0$. Thus, by Remark 3.1, condition **B (a)** holds, with the corresponding limiting distribution function $Q_{0,11}(t)$. According to condition **H**, condition **B (b)** holds for the distribution function $Q_{0,11}(t)$. Analogously, in relation (3.48), either expectation $e_{\varepsilon,12} = 0$, if $p_{\varepsilon,12} = 0$, for $\varepsilon \in [0, 1]$, or $e_{\varepsilon,12} \to 0$ as $\varepsilon \to 0$, if $0 < p_{\varepsilon,12} \to 0$ as $\varepsilon \to 0$. It follows from this remark and conditions **G–I** that the expectations $\hat{e}_{\varepsilon,11} \to \hat{e}_{0,11} = e_{0,11}$ as $\varepsilon \to 0$. Note also that condition $\mathbf{M}_3$ **(a)** implies that expectation $e_{0,11} = e_{0,1}$. Thus, condition **C** holds, with the corresponding limiting expectation $e_{0,11} = e_{0,1}$. Relation (3.46) and condition **J** imply that condition **D** holds. As was mentioned above, in this case, $f_\varepsilon \equiv 0$. Thus, all conditions all conditions of Theorem 3.1 hold, and the ergodic relation given in this theorem takes place for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$. In this case, it takes the form of relation (3.43), where one should choose $i, j = 1$.

Second, let us analyse the asymptotic behaviour of probabilities $P_{\varepsilon,12}(t, A)$.

Holding of conditions **A–C** was pointed above.

If $\eta_\varepsilon(0) = 1$, then $\eta_\varepsilon(t) = 1$ for $t \in [0, \tau_{\varepsilon,1})$, and $\hat{\tau}_{\varepsilon,1} = \tau_{\varepsilon,1}$, if $\eta_{\varepsilon,1} = 1$. Also, $\eta_\varepsilon(t) = 2$, for $t \in [\tau_{\varepsilon,1}, \hat{\tau}_{\varepsilon,1})$, if $\eta_{\varepsilon,1} = 2$. Therefore, for every $A \in \mathcal{B}_{\mathbb{X}}$, $t \geq 0$,

$$\hat{q}_{\varepsilon,12}(t, A) = \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 2, \hat{\tau}_{\varepsilon,1} > t\}$$
$$\leq \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 2, \hat{\tau}_{\varepsilon,1} > t, \tau_{\varepsilon,1} \leq t, \eta_{\varepsilon,1} = 2\}$$
$$\leq \mathsf{P}_1\{\tau_{\varepsilon,1} \leq t, \eta_{\varepsilon,1} = 2\} \leq p_{\varepsilon,12}. \tag{3.49}$$

Since, $p_{\varepsilon,12} = 0$, for $\varepsilon \in [0, 1]$ or $p_{\varepsilon,12} \to 0$ as $\varepsilon \to 0$, condition **D** holds for function $\hat{q}_{\varepsilon,12}(t, A)$ with the corresponding limiting function $\hat{q}_{0,12}(t, A) = 0$, $t \geq 0$, for every $A \in \mathcal{B}_{\mathbb{X}}$. Thus, all conditions of Theorem 3.1 hold, and the ergodic relation given in this theorem takes place for probabilities $P_{\varepsilon,12}(t_\varepsilon, A)$. In this case, it takes the form of relation (3.43), where one should choose $i = 1$, $j = 2$.

Third, let us analyse asymptotic behaviour of probabilities $P_{\varepsilon,2j}(t, A)$, $j = 1, 2$. In this case, we do prefer to consider $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \geq 0$ as the regenerative process with transition period $[0, \hat{\tau}'_{\varepsilon,1})$ and regenerative times $\hat{\tau}'_{\varepsilon,0}$, $\hat{\tau}'_{\varepsilon,1} = \tilde{\tau}_{\varepsilon,1}$, $\hat{\tau}'_{\varepsilon,2}$, $\hat{\tau}'_{\varepsilon,3}$, ….

The shifted process $(\xi_\varepsilon(\hat{\tau}'_{\varepsilon,1}+t), \eta_\varepsilon(\hat{\tau}'_{\varepsilon,1}+t)), t \geq 0$ is a standard regenerative process. If $\eta_\varepsilon(0) = 2$, then $\eta_\varepsilon(\tilde{\tau}_{\varepsilon,1}) = 1$. That is why, probabilities $P_{\varepsilon,1j}(t, A)$ play for this process the role of probabilities $P_\varepsilon^{(1)}(t, A)$ defined out in Sect. 3.2.

The distribution function $\mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \leq t\} = \tilde{Q}_{\varepsilon,21}(t)$ and the Laplace transform $\tilde{\phi}_{\varepsilon,12}(s) = \int_0^\infty e^{-st}\tilde{Q}_{\varepsilon,12}(dt)$ for the duration of the transition period $[0, \tilde{\tau}_{\varepsilon,1})$ are given, respectively in relations (3.39) and (3.40). These relations and conditions **G**, **H** and $\mathbf{M}_3$ **(a)** imply that Laplace transforms $\tilde{\phi}_{\varepsilon,12}(s) \to \tilde{\phi}_{0,12}(s)$ as $\varepsilon \to 0$, for $s \geq 0$. Thus, by Remark 3.1, condition **E** holds. All conditions of Theorem 3.2 hold, and the corresponding ergodic relation for probabilities $P_{\varepsilon,1j}(t_\varepsilon, A)$, $j = 1, 2$ also holds for probabilities $P_{\varepsilon,2j}(t_\varepsilon, A)$, $j = 1, 2$.

Due to symmetricity of conditions **F**–**J** with respect to the indices $i, j = 1, 2$ the corresponding asymptotic analysis for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$, $i, j = 1, 2$ (under assumption of holding condition $\mathbf{M}_3$**(b)**) is analogous to the above asymptotic analysis for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$, $i.j = 1, 2$ (under assumption of holding condition $\mathbf{M}_3$**(a)**). The corresponding ergodic relation (3.44) takes place for the above probabilities, under assumption of holding condition $\mathbf{M}_3$**(b)**. □

3.4 Super-Long and Long Time Ergodic Theorems for Singularly Perturbed Alternating Regenerative Processes

In this section, we present super-long and long time individual ergodic theorems for singularly perturbed alternating regenerative processes. We also present in this section the special procedure of time scaling for perturbed regenerative processes. It is essentially used in the corresponding proofs.

3.4.1 Time Scaling for Perturbed Regenerative Processes

Let return back to the model of perturbed regenerative processes with regenerative lifetimes introduced in Sect. 3.2.1. So, let $\xi_\varepsilon(t), t \geq 0$ be, for every $\varepsilon \in [0, 1]$, a regeneration process with regeneration times $\tau_{\varepsilon,n}, n = 0, 1, \ldots$ and a regenerative lifetime μ_ε constructed using the triplets $\langle\bar{\xi}_{\varepsilon,n} = \langle\xi_{\varepsilon,n}(t), t \geq 0\rangle, \kappa_{\varepsilon,n}, \mu_{\varepsilon,n}\rangle$ introduced in Sect. 3.2.1.

Let also $v_\varepsilon, \varepsilon \in (0, 1]$ be a positive function. We also choose some $v_0 \in [0, \infty]$.

In some cases, it can be useful to replace, for every $\varepsilon \in (0, 1]$, the above triplet by new one, $\langle\bar{\xi}_{\varepsilon,v_\varepsilon,n}=\langle\xi_{\varepsilon,v_\varepsilon,n}(t)=\xi_{\varepsilon,n}(tv_\varepsilon), t \geq 0\rangle, \kappa_{\varepsilon,v_\varepsilon,n}=v_\varepsilon^{-1}\kappa_{\varepsilon,n}, \mu_{\varepsilon,v_\varepsilon,n} = v_\varepsilon^{-1}\mu_{\varepsilon,n}\rangle$.

Respectively, the above regenerative process $\xi_\varepsilon(t), t \geq 0$ will be, for every $\varepsilon \in (0, 1]$, transformed in the new process $\xi_{\varepsilon,v_\varepsilon}(t) = \xi_\varepsilon(tv_\varepsilon), t \geq 0$. Obviously. $\xi_{\varepsilon,v_\varepsilon}(t), t \geq 0$ is also a regenerative process, with new regenerative times $\tau_{\varepsilon,v_\varepsilon,n} = v_\varepsilon^{-1}\tau_{\varepsilon,n}, n = 0, 1, \ldots$ and new lifetime $\mu_{\varepsilon,v_\varepsilon} = v_\varepsilon^{-1}\mu_\varepsilon$.

We also should introduce a limiting triplet $\langle \bar{\xi}_{0,v_0,n} = \langle \xi_{0,v_0,n}(t),\ t \geq 0\rangle, \kappa_{0,v_0,n}, \mu_{0,v_0,n}\rangle$, which possess the corresponding properties described in Sect. 3.2.1, and the corresponding limiting regenerative process $\xi_{0,v_0}(t) = \xi_{0,v_0,n}(t - \tau_{0,v_0,n-1})$ for $t \in [\tau_{0,v_0,n-1}, \tau_{0,v_0,n})$, $n = 1, 2, \ldots$, regeneration times $\tau_{0,v_0,n} = \kappa_{0,v_0,1} + \cdots + \kappa_{0,v_0,n}$, $n = 1, 2, \ldots, \tau_{0,v_0,0} = 0$, and a regenerative lifetime, $\mu_{0,v_0} = \kappa_{0,v_0,1} + \cdots + \kappa_{0,v_0,v_{0,v_0}-1} + \mu_{0,v_0,v_{0,v_0}}$, where $v_{0,v_0} = \min(n \geq 1 : \mu_{0,v_0,n} < \kappa_{0,v_0,n})$.

In such model, we can assume the the corresponding conditions **A**–**D** hold for the transformed regenerative processes $\xi_{\varepsilon,v_\varepsilon}(t)$, $t \geq 0$, their regeneration times $\tau_{\varepsilon,v_\varepsilon,n}$, $n = 0, 1, \ldots$ and lifetimes $\mu_{\varepsilon,v_\varepsilon}$.

It worth to note that the probabilities $P_{\varepsilon,v_\varepsilon}(t, A) = \mathsf{P}\{\xi_{\varepsilon,v_\varepsilon}(t) \in A, \mu_{\varepsilon,v_\varepsilon} > t\} = P_\varepsilon(tv_\varepsilon, A) = \mathsf{P}\{\xi_\varepsilon(tv_\varepsilon) \in A, \mu_\varepsilon > tv_\varepsilon\}$, $t \geq 0$, for $\varepsilon \in (0, 1]$.

The basic renewal equation (3.7) for probabilities $P_{\varepsilon,v_\varepsilon}(t, A)$ takes, for $\varepsilon \in (0, 1]$, the following form, for $A \in \mathcal{B}_{\mathbb{X}}$,

$$P_{\varepsilon,v_\varepsilon}(t, A) = q_{\varepsilon,v_\varepsilon}(t, A) + \int_0^t P_{\varepsilon,v_\varepsilon}(t - s, A) F_{\varepsilon,v_\varepsilon}(ds),\ t \geq 0, \tag{3.50}$$

where $q_{\varepsilon,v_\varepsilon}(t, A) = \mathsf{P}\{\xi_{\varepsilon,v_\varepsilon}(t) \in A, \tau_{\varepsilon,v_\varepsilon,1} \wedge \mu_{\varepsilon,v_\varepsilon} > t\} = q_\varepsilon(tv_\varepsilon, A) = \mathsf{P}\{\xi_\varepsilon(tv_\varepsilon) \in A, \tau_{\varepsilon,1} \wedge \mu_\varepsilon > tv_\varepsilon\}$ and $F_{\varepsilon,v_\varepsilon}(t) = \mathsf{P}\{\tau_{\varepsilon,v_\varepsilon,1} \leq t, \mu_{\varepsilon,v_\varepsilon} \geq \tau_{\varepsilon,v_\varepsilon,1}\} = \mathsf{P}\{\tau_{\varepsilon,1} \leq tv_\varepsilon, v_\varepsilon^{-1} \mu_\varepsilon \geq v_\varepsilon^{-1}\tau_{\varepsilon,1}\}$.

We shall see in the next section that the above scaling of time transformation can be effectively used in ergodic theorems for singularly perturbed alternating regenerative processes, where aggregated regeneration times can be stochastically unbounded as $\varepsilon \to 0$. In such models, we shall use time scaling factors $0 < v_\varepsilon \to v_0 = \infty$ as $\varepsilon \to 0$, and refer to v_ε as to time compression factors.

3.4.2 *Singularly Perturbed Alternating Regenerative Processes*

Let us now consider the alternating regenerative processes with the singular perturbation model, where additionally to **F**–**J**, the following condition holds:

$\mathbf{N}_1$: $0 < p_{\varepsilon,12} \to p_{0,12} = 0$ as $\varepsilon \to 0$ and $0 < p_{\varepsilon,21} \to p_{0,21} = 0$ as $\varepsilon \to 0$.

The case, where condition $\mathbf{N}_1$ holds, is the most interesting. Here, we should also assume that probabilities $p_{\varepsilon,12}$ and $p_{\varepsilon,21}$ are asymptotically comparable in the sense that the condition $\mathbf{K}_\beta$ holds for some $\beta \in [0, \infty]$.

Let us define function $v_\varepsilon = p_{\varepsilon,12}^{-1} + p_{\varepsilon,21}^{-1}$. Obviously, $0 < v_\varepsilon \to v_0 = \infty$ as $\varepsilon \to 0$. Also, $p_{\varepsilon,12}^{-1}/v_\varepsilon \to (1 + \beta)^{-1}$ and $p_{\varepsilon,21}^{-1}/v_\varepsilon \to (1 + \beta^{-1})^{-1}$ as $\varepsilon \to 0$.

As was pointed out in Sect. 3.4.1, process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$ is a regenerative process with regeneration times $\hat{\tau}_{\varepsilon,n}$, $n = 0, 1, \ldots$. It is also a regenerative process with the transition period $[0, \hat{\tau}'_{\varepsilon,1})$ and regeneration times $\hat{\tau}'_{\varepsilon,n}$, $n = 0, 1, \ldots$.

Unfortunately, the model with aggregated regeneration times $\hat{\tau}_{\varepsilon,n}$ does not work in this case. Indeed, conditions **G**, **H** and $\mathbf{N}_1$ implies that $\phi_{\varepsilon,12}(s) \to \phi_{0,12}(s) = 0$ as $\varepsilon \to 0$, for $s \geq 0$ and, thus, using relation (3.36), we get $\hat{\phi}_{\varepsilon,11}(s) = \phi_{\varepsilon,11}(s) + \phi_{\varepsilon,12}(s)\frac{1}{1-\phi_{\varepsilon,22}(s)}\phi_{\varepsilon,21}(s) \to \phi_{0,11}(s)$ as $\varepsilon \to 0$, for $s \geq 0$. Thus, the distributions of regeneration times $\hat{Q}_{\varepsilon,11}(\cdot) \Rightarrow \hat{Q}_{0,11}(\cdot) = Q_{0,11}(\cdot)$ as $\varepsilon \to 0$. At the same time, conditions **G**–**I**, $\mathbf{N}_1$ and relation (3.37) imply that, in this case, $\hat{e}_{\varepsilon,11} = \frac{e_{\varepsilon,1}p_{\varepsilon,21}+e_{\varepsilon,2}p_{\varepsilon,12}}{p_{\varepsilon,21}} \to e_{0,1} + e_{0,2}\beta = e_{0,11} + e_{0,22}\beta$ as $\varepsilon \to 0$. This makes it impossible to use Theorems 3.1–3.3, which require convergence of expectations for regeneration times to the first moment of the corresponding limiting distribution for regeneration times. In the above case, $\hat{e}_{0,11} = e_{0,11} \neq e_{0,11} + e_{0,22}\beta$, if $\beta > 0$.

Fortunately, we can use an alternative model with aggregated regeneration times $\check{\tau}_{\varepsilon,n}$ introduced in Sect. 3.2.7. Process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$ is a regenerative process with regeneration times $\check{\tau}_{\varepsilon,n}$, $n = 0, 1, \ldots$. It is also a regenerative process with the transition period $[0, \check{\tau}'_{\varepsilon,1})$ and regeneration times $\check{\tau}'_{\varepsilon,n}$, $n = 0, 1, \ldots$.

Let us analyse the asymptotic behaviour for probabilities $P_{\varepsilon,11}(t, A)$. In this case, we do prefer to consider $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \geq 0$ as the standard regenerative process with regeneration times $\check{\tau}_{\varepsilon,n}$, $n = 0, 1, \ldots$.

The renewal equation (3.7) for probabilities $P_{\varepsilon,11}(t, A)$ takes, in this case, the following form,

$$P_{\varepsilon,11}(t, A) = \check{q}_{\varepsilon,1}(t, A) + \int_0^t P_{\varepsilon,11}(t - s, A)\check{Q}_{\varepsilon,11}(ds), t \geq 0, \tag{3.51}$$

where $\check{q}_{\varepsilon,1}(t, A) = \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \check{\tau}_{\varepsilon,1} > t\}$, $t \geq 0$ and $\check{Q}_{\varepsilon,11}(t) = \mathsf{P}_1\{\check{\tau}_{\varepsilon,1} \leq t\}$, $t \geq 0$.

If $\eta_\varepsilon(0) = 1$, then $\eta_\varepsilon(t) = 1$ for $t \in [0, \tilde{\tau}_{\varepsilon,1})$, and $\eta_\varepsilon(t) = 2$, for $t \in [\tilde{\tau}_{\varepsilon,1}, \check{\tau}_{\varepsilon,1})$. Therefore, for every $A \in \mathcal{B}_{\mathbb{X}}$, $t \geq 0$,

$$\begin{aligned}\check{q}_{\varepsilon,1}(t, A) &= \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \check{\tau}_{\varepsilon,1} > t\} \\ &= \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \tilde{\tau}_{\varepsilon,1} > t\} = \tilde{q}_{\varepsilon,1}(t, A).\end{aligned} \tag{3.52}$$

In this case, $\check{Q}_{\varepsilon,11}(t)$ is the distribution function of the first return time to state 1 after first hitting to state 2, for the semi-Markov process $\eta_\varepsilon(t)$. It can be expressed in terms of convolutions of transition probabilities for this semi-Markov process,

$$\check{Q}_{\varepsilon,11}(t) = \tilde{Q}_{\varepsilon,12}(t) * \tilde{Q}_{\varepsilon,21}(t), t \geq 0. \tag{3.53}$$

where, for $i, j \in \mathbb{Y}$, $i \neq j$,

$$\tilde{Q}_{\varepsilon,ij}(t) = \sum_{n=0}^{\infty} Q^{*n}_{\varepsilon,ii}(t) * Q_{\varepsilon,ij}(t), \ t \geq 0, \tag{3.54}$$

According relation (3.53), the distribution function $\check{Q}_{\varepsilon,11}(t)$ of return time $\check{\tau}_{\varepsilon,1}$ is the convolution of two distribution functions, $\tilde{Q}_{\varepsilon,12}(t)$ and $\tilde{Q}_{\varepsilon,21}(t)$. This means that

return time $\check{\tau}_{\varepsilon,1}$ is the sum of two independent random variables $\tilde{\tau}_{\varepsilon,1}$ and $\check{\tau}_{\varepsilon,1} - \tilde{\tau}_{\varepsilon,1}$, which have the distribution functions, respectively, $\tilde{Q}_{\varepsilon,12}(t)$ and $\tilde{Q}_{\varepsilon,21}(t)$. The former one is the distribution of the first hitting time of state 2 from state 1, the latter one is the distribution of the first hitting time of state 1 from state 2, for the semi-Markov process $\eta_\varepsilon(t)$.

Remind that we assume, $\eta_\varepsilon(0) = \eta_\varepsilon = 1$. In this case, (a) the return time $\tilde{\tau}_{\varepsilon,1}$ is a random sum, $\tilde{\tau}_{\varepsilon,1} = \sum_{n=1}^{\theta_\varepsilon[0]} \kappa_{\varepsilon,1,n}$, where (b) the random index, $\theta_\varepsilon[0] = \min(n \geq 1 : \eta_{\varepsilon,1,n} = 1)$ has the geometric distribution with parameter $p_{\varepsilon,12}$, i.e., it takes value n with probability $p_{\varepsilon,11}^{n-1} p_{\varepsilon,12}$, for $n = 1, 2, \ldots$.

Relation (b) and condition $\mathbf{N}_1$ imply that random variables,

$$p_{\varepsilon,12}\theta_\varepsilon[0] \xrightarrow{d} \zeta \text{ as } \varepsilon \to 0, \tag{3.55}$$

where ζ is a random variable exponentially distributed, with parameter 1.

Random variables $\kappa_{\varepsilon,1,n}$, $n = 1, 2, \ldots$ are i.i.d. random variables with the distribution function $F_{\varepsilon,1}(t) = \mathsf{P}_1\{\kappa_{\varepsilon,1,1} \leq t\} = Q_{\varepsilon,11}(t) + Q_{\varepsilon,12}(t)$, $t \geq 0$. Conditions **H** and **I** imply that (c) distributions $F_{\varepsilon,1}(\cdot) \Rightarrow F_{0,1}(\cdot)$ as $\varepsilon \to 0$ and (d) expectations $e_{\varepsilon,1} = \mathsf{E}_1\kappa_{\varepsilon,1,1} = \int_0^\infty s F_{\varepsilon,1}(ds) \to e_{0,1} = \int_0^\infty s F_{0,1}(ds)$ as $\varepsilon \to 0$.

Relations (c) and (d) imply that, for any integer-valued function $0 \leq n_\varepsilon \to \infty$ as $\varepsilon \to 0$,

$$n_\varepsilon^{-1} \sum_{k=1}^{n_\varepsilon} \kappa_{\varepsilon,1,n} \xrightarrow{d} e_{0,1} \text{ as } \varepsilon \to 0. \tag{3.56}$$

Indeed, let $0 < s_k \to \infty$ as $k \to \infty$ be a sequence of continuity points for the distribution function $F_{0,1}(t)$. The above relations (c) and (d) obviously imply that, for any $t > 0$,

$$\begin{aligned}\overline{\lim_{\varepsilon\to 0}} \int_{tn_\varepsilon}^\infty s F_{\varepsilon,1}(ds) &\leq \overline{\lim_{\varepsilon\to 0}} \int_{s_k}^\infty s F_{\varepsilon,1}(ds) = \overline{\lim_{\varepsilon\to 0}}(e_{\varepsilon,1} - \int_0^{s_k} s F_{\varepsilon,1}(ds)) \\ &= e_{0,1} - \int_0^{s_k} s F_{0,1}(ds) \to 0 \text{ as } k \to \infty,\end{aligned} \tag{3.57}$$

and, thus, the following relation holds, for any $t > 0$,

$$\overline{\lim_{\varepsilon\to 0}} \int_{tn_\varepsilon}^\infty s F_{\varepsilon,1}(ds) = 0. \tag{3.58}$$

Relation (3.58) implies that, for any $t > 0$,

$$\begin{aligned}n_\varepsilon \mathsf{P}_1\{n_\varepsilon^{-1}\kappa_{\varepsilon,1,1} > t\} &= n_\varepsilon(1 - F_{\varepsilon,1}(tn_\varepsilon)) \\ &\leq t^{-1} \int_{tn_\varepsilon}^\infty s F_{\varepsilon,1}(ds) \to 0 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.59}$$

Also, relations (d) and (3.58) implies that, for any $t > 0$,

$$n_\varepsilon \mathsf{E}_1 n_\varepsilon^{-1} \kappa_{\varepsilon,1,1} \mathrm{I}(n_\varepsilon^{-1} \kappa_{\varepsilon,1,1} \le t) = \int_0^{tn_\varepsilon} s dF_{\varepsilon,1}(ds) \to e_{0,1} \text{ as } \varepsilon \to 0. \tag{3.60}$$

Relations (3.59) and (3.60) imply, by the criterion of central convergence, that relation (3.56) holds.

The random index $\theta_\varepsilon[0]$ and the random variables $\kappa_{\varepsilon,1,n}$, $n = 1, 2, \ldots$ are dependent. Nevertheless, since the limit in relation (3.56) is non-random, relations (3.55) and (3.56) imply that stochastic processes,

$$(p_{\varepsilon,12}\theta_\varepsilon[0], \sum_{n \le t p_{\varepsilon,12}^{-1}} \kappa_{\varepsilon,1,n}, t \ge 0 \xrightarrow{d} (\zeta, te_{0,1}), t \ge 0 \text{ as } \varepsilon \to 0. \tag{3.61}$$

By well known results about convergence of randomly stopped stochastic processes (for example, Theorem 2.2.1 [52]), representation (a) and relation (3.61) imply that random variables,

$$p_{\varepsilon,12}\tilde{\tau}_{\varepsilon,1} \xrightarrow{d} e_{0,1}\zeta \text{ as } \varepsilon \to 0. \tag{3.62}$$

Since, $p_{\varepsilon,12}^{-1}/v_\varepsilon \to (1+\beta)^{-1}$ as $\varepsilon \to 0$, relation (k) implies the following relation,

$$\tilde{Q}_{\varepsilon,v_\varepsilon,12}(\cdot) \Rightarrow \tilde{Q}_{0,v_0,12}(\cdot) \text{ as } \varepsilon \to 0, \tag{3.63}$$

where $\tilde{Q}_{0,v_0,12}(t) = \mathsf{P}\{e_{0,1}\frac{1}{1+\beta}\zeta \le t\}$, $t \ge 0$ is the distribution function of an exponentially distributed random variable, with parameter $e_{0,1}^{-1}(1+\beta)$.

Since the random variable $\check{\tau}_{\varepsilon,1} - \tilde{\tau}_{\varepsilon,1}$ has distribution function $\tilde{Q}_{\varepsilon,21}(t)$, one can, in the way absolutely analogous with relation (3.62), prove the following relation,

$$p_{\varepsilon,21}(\check{\tau}_{\varepsilon,1} - \tilde{\tau}_{\varepsilon,1}) \xrightarrow{d} e_{0,2}\zeta \text{ as } \varepsilon \to 0, \tag{3.64}$$

and, in sequel,

$$\tilde{Q}_{\varepsilon,v_\varepsilon,21}(\cdot) \Rightarrow \tilde{Q}_{0,v_0,21}(\cdot) \text{ as } \varepsilon \to 0, \tag{3.65}$$

where $\tilde{Q}_{0,v_0,21}(t) = \mathsf{P}\{e_{0,2}\frac{1}{1+\beta^{-1}}\zeta \le t\}$, $t \ge 0$ is the distribution function of an exponentially distributed random variable, with parameter $e_{0,2}^{-1}(1+\beta^{-1})$.

Let $\check{Q}_{\varepsilon,v_\varepsilon,11}(t) = \check{Q}_{\varepsilon,11}(tv_\varepsilon) = \mathsf{P}_1\{\check{\tau}_{\varepsilon,1}/v_\varepsilon \le t\}$, $t \ge 0$ be the distribution function of the normalised return time $v_\varepsilon^{-1}\check{\tau}_{\varepsilon,1}$.

Relations (3.53), (3.63) and (3.65) imply that,

$$\check{Q}_{\varepsilon,v_\varepsilon,11}(\cdot) \Rightarrow \check{Q}_{0,v_0,11}(\cdot) \text{ as } \varepsilon \to 0, \tag{3.66}$$

where $\check{Q}_{0,v_0,11}(t) = \mathsf{P}\{e_{0,1}\frac{1}{1+\beta}\zeta_1 + e_{0,2}\frac{1}{1+\beta^{-1}}\zeta_2 \le t\}, t \ge 0$ is the distribution function of the linear combination of two independent random variables ζ_1 and ζ_2, exponentially distributed, with parameter 1.

Note that in that cases $\beta = 0$ or $\beta = \infty$, respectively, the second or the first random variable in the above sum vanishes in zero. In this case, $\check{Q}_{0,v_0,11}(t)$ is an exponential distribution function with parameter, respectively, $e_{0,1}^{-1}$ or $e_{0,2}^{-1}$.

Also, that above representation (a) for the random variable $\tilde{\tau}_{\varepsilon,1}$, as the random sum, implies that,

$$\begin{aligned}
\tilde{e}_{\varepsilon,v_\varepsilon,12} &= \int_0^\infty s\,\tilde{Q}_{\varepsilon,v_\varepsilon,12}(ds) = v_\varepsilon^{-1}\mathsf{E}\sum_{n=1}^{\theta_\varepsilon[0]}\kappa_{\varepsilon,1,n} \\
&= v_\varepsilon^{-1}\mathsf{E}\sum_{n=1}^{\infty}\kappa_{\varepsilon,1,n}\mathrm{I}(\theta_\varepsilon[0] > n-1) \\
&= v_\varepsilon^{-1}\mathsf{E}\sum_{n=1}^{\infty}\kappa_{\varepsilon,1,n}\mathrm{I}(\eta_{\varepsilon,1,k} = 1, 1 \le k \le n-1) \\
&= v_\varepsilon^{-1}\sum_{n=1}^{\infty}\mathsf{E}\kappa_{\varepsilon,1,n}\mathsf{E}\mathrm{I}(\eta_{\varepsilon,1,k} = 1, 1 \le k \le n-1) \\
&= v_\varepsilon^{-1}\sum_{n=1}^{\infty} e_{\varepsilon,1}\,p_{\varepsilon,11}^{n-1} = \frac{e_{\varepsilon,1}}{v_\varepsilon\,p_{\varepsilon,12}}.
\end{aligned} \tag{3.67}$$

Analogous formula also takes place,

$$\tilde{e}_{\varepsilon,v_\varepsilon,21} = v_\varepsilon^{-1}\mathsf{E}(\check{\tau}_{\varepsilon,1} - \tilde{\tau}_{\varepsilon,1}) = \frac{e_{\varepsilon,2}}{v_\varepsilon\,p_{\varepsilon,21}}. \tag{3.68}$$

Relations (3.67) and (3.68) imply the following relation,

$$\begin{aligned}
\check{e}_{\varepsilon,v_\varepsilon,11} &= \tilde{e}_{\varepsilon,v_\varepsilon,12} + \tilde{e}_{\varepsilon,v_\varepsilon,21} = e_{\varepsilon,1}\frac{1}{v_\varepsilon\,p_{\varepsilon,12}} + e_{\varepsilon,2}\frac{1}{v_\varepsilon\,p_{\varepsilon,21}} \\
&\to e_{0,1}\frac{1}{1+\beta} + e_{0,2}\frac{1}{1+\beta^{-1}} = \check{e}_{0,v_0,11} = \int_0^\infty s\,\check{Q}_{0,v_0,11}(ds).
\end{aligned} \tag{3.69}$$

The above remarks prompt us how to apply the scaling of time transformation with compression function v_ε, described in Sect. 3.4.1, to the regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t)), t \ge 0$ with regeneration times $\check{\tau}_{\varepsilon,n}, n = 0, 1, \ldots$.

So, let us consider, for every $\varepsilon \in (0, 1]$, the compressed in time version of the regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t)), t \ge 0$ with regeneration times $\check{\tau}_{\varepsilon,n}, n = 0, 1, \ldots$. It is the regenerative process $(\xi_{\varepsilon,v_\varepsilon}(t), \eta_{\varepsilon,v_\varepsilon}(t)), t \ge 0 = (\xi_\varepsilon(tv_\varepsilon), \eta_\varepsilon(tv_\varepsilon)), t \ge 0$ with regeneration times $\tau_{\varepsilon,v_\varepsilon,n} = v_\varepsilon^{-1}\check{\tau}_{\varepsilon,n}, n = 0, 1, \ldots$.

The renewal type Eq. (3.7) takes for probabilities $P_{\varepsilon,v_\varepsilon,11}(t, A) = P_{\varepsilon,11}(tv_\varepsilon, A)$ the following form,

$$P_{\varepsilon,v_\varepsilon,11}(t, A) = \check{q}_{\varepsilon,v_\varepsilon,1}(t, A) + \int_0^t P_{\varepsilon,v_\varepsilon,11}(t-s, A)\check{Q}_{\varepsilon,v_\varepsilon,11}(ds), t \geq 0, \quad (3.70)$$

where $\check{q}_{\varepsilon,v_\varepsilon,1}(t, A) = \mathsf{P}_1\{\xi_{\varepsilon,v_\varepsilon,}(t) \in A, \eta_{\varepsilon,v_\varepsilon}(t) = 1, \check{\tau}_{\varepsilon,v_\varepsilon,1} > t\} = \check{q}_{\varepsilon,1}(tv_\varepsilon, A) = \mathsf{P}_1\{\xi_\varepsilon(tv_\varepsilon) \in A, \eta_\varepsilon(tv_\varepsilon) = 1, v_\varepsilon^{-1}\check{\tau}_{\varepsilon,1} > t\}, t \geq 0$ and $\check{Q}_{\varepsilon,v_\varepsilon,11}(t) = \mathsf{P}_1\{\check{\tau}_{\varepsilon,v_\varepsilon,1} \leq t\} = \check{Q}_{\varepsilon,11}(tv_\varepsilon) = \mathsf{P}_1\{v_\varepsilon^{-1}\check{\tau}_{\varepsilon,1} \leq t\}, t \geq 0.$

The corresponding limiting regenerative process $(\xi_{0,v_0}(t), \eta_{0,v_0}(t))$, $t \geq 0$ and the regeneration times $\check{\tau}_{0,v_0,n}, n = 0, 1, \ldots$ will be defined in the next subsection, after computing the corresponding limits for functions $\check{q}_{\varepsilon,v_\varepsilon,1}(t, A)$ and distribution functions $\check{Q}_{\varepsilon,v_\varepsilon,11}(t)$.

3.4.3 Locally Uniform Convergence of Functions and Convergence of Lebesgue Integrals in the Scheme of Series

In this subsection, we formulate two useful propositions concerned locally uniform convergence of functions and convergence of Lebesgue integrals in the scheme of series. The proofs can be found, for example, in book [14]. We slightly modify these propositions for the case, where the corresponding functions and measures are defined on a half-line.

Let $f_\varepsilon(s)$ be, for every $\varepsilon \in [0, 1]$, a real-valued bounded Borel functions defined on $\mathbb{R}_+ = [0, \infty)$. We use the symbol $f_\varepsilon(s) \xrightarrow{\text{U}} f_0(s)$ as $\varepsilon \to 0$ to indicate that functions $f_\varepsilon(\cdot)$ converge to function $f_0(\cdot)$ locally uniformly at a point $s \in [0, \infty)$ as $\varepsilon \to 0$. This means that,

$$\lim_{0<u\to 0} \overline{\lim_{\varepsilon\to 0}} \sup_{-(u\wedge s)\leq v\leq u} |f_\varepsilon(s+v) - f_0(s)| = 0. \quad (3.71)$$

Lemma 3.1 *Functions* $f_\varepsilon(s) \xrightarrow{\text{U}} f_\varepsilon(s)$ *as* $\varepsilon \to 0$ *if and only if* $(\boldsymbol{\alpha})$ $f_\varepsilon(s_\varepsilon) \to f_0(s)$ *as* $\varepsilon \to 0$, *for any* $0 \leq s_\varepsilon \to s$ *as* $\varepsilon \to 0$.

Let $\mathcal{B}_+$ denote the Borel σ-algebra on $\mathbb{R}_+$ and let $\mu_\varepsilon(A)$ be, for every $\varepsilon \in [0, 1]$, a finite measure on $\mathcal{B}_+$. We use the symbol $\mu_\varepsilon(\cdot) \Rightarrow \mu_0(\cdot)$ as $\varepsilon \to 0$ to indicate that the measures $\mu_\varepsilon(A)$ weakly converge to a measure $\mu_0(A)$ as $\varepsilon \to 0$. This means that, for all $0 \leq v < \infty$ such that the limiting measure has not an atom in the point v,

$$\mu_\varepsilon([0, v]) \to \mu_0([0, v]) \text{ as } \varepsilon \to 0. \quad (3.72)$$

Lemma 3.2 *Let the following conditions hold*: $(\boldsymbol{\alpha})$ $\mu_\varepsilon(\cdot) \Rightarrow \mu_0(\cdot)$ *as* $\varepsilon \to 0$; $(\boldsymbol{\beta})$ $\mu_\varepsilon(\mathbb{R}_+) \to \mu_0(\mathbb{R}_+)$ *as* $\varepsilon \to 0$; $(\boldsymbol{\gamma})$ $\overline{\lim}_{\varepsilon\to 0} \sup_{s\in\mathbb{R}_+} |f_\varepsilon(s)| < \infty$; $(\boldsymbol{\delta})$ $f_\varepsilon(s) \xrightarrow{\mathrm{U}} f_0(s)$ *as* $\varepsilon \to 0$, *for* $s \in S$, *where* S *is some subset of* $\mathcal{B}_+$ *such that* $\mu_0(\overline{S}) = 0$. *Then,*

$$\int_{\mathbb{R}_+} f_\varepsilon(s)\mu_\varepsilon(ds) \to \int_{\mathbb{R}_+} f_0(s)\mu_0(ds) \text{ as } \varepsilon \to 0. \tag{3.73}$$

3.4.4 Super-Long Time Ergodic Theorems for Singularly Perturbed Alternating Regenerative Processes

In this subsection, we describe the asymptotic behaviour for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ for so-called "super-long" times $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ satisfying the following relation,

$$t_\varepsilon/v_\varepsilon \to \infty \text{ as } \varepsilon \to 0. \tag{3.74}$$

The corresponding limits for stationary probabilities for perturbed semi-Markov processes $\eta_{\varepsilon,v_\varepsilon}(t)$ take, for $\beta \in [0, \infty]$, the following form,

$$\rho_1(\beta) = e_{0,1}\alpha_1(\beta)/e(\beta),\ \rho_2(\beta) = e_{0,2}\alpha_2(\beta)/e(\beta), \tag{3.75}$$

where

$$\alpha_1(\beta) = (1+\beta)^{-1},\ \alpha_2(\beta) = (1+\beta^{-1})^{-1}, \tag{3.76}$$

and

$$e(\beta) = e_{0,1}\alpha_1(\beta) + e_{0,2}\alpha_2(\beta). \tag{3.77}$$

Note that $\rho_1(\beta), \rho_2(\beta) \in (0, 1)$, if $\beta \in (0, \infty)$, while $\rho_1(\beta) = 1, \rho_2(\beta) = 0$, if $\beta = 0$, and $\rho_1(\beta) = 0, \rho_2(\beta) = 1$, if $\beta = \infty$.

The corresponding limiting probabilities for singularly perturbed alternating regenerative processes take the following form,

$$\pi^{(\beta)}_{0,j}(A) = \rho_j(\beta)\pi_{0,j}(A),\ A \in \mathcal{B}_{\mathbb{X}},\ j = 1, 2. \tag{3.78}$$

It is useful to note that the above limiting probabilities coincide with the corresponding limiting probabilities for regularly perturbed alternating regenerative processes with parameter $\beta = p_{0,12}/p_{0,21}$ given in relations (3.31), (3.41), and (3.42).

The following theorem takes place.

Theorem 3.8 *Let conditions* **F**–**J**, $\mathbf{N}_1$ *hold and, also, condition* $\mathbf{K}_\beta$ *holds for some* $\beta \in [0, \infty]$. *Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon/v_\varepsilon \to \infty$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi^{(\beta)}_{0,j}(A) \text{ as } \varepsilon \to 0. \tag{3.79}$$

Proof We are going to prove that all conditions of Theorem 3.1 hold for the regenerative processes $(\xi_{\varepsilon,v_\varepsilon}(t), \eta_{\varepsilon,v_\varepsilon}(t)) = (\xi_\varepsilon(tv_\varepsilon), \eta_\varepsilon(tv_\varepsilon))$, $t \geq 0$ with regeneration times $\check{\tau}_{\varepsilon,v_\varepsilon,n} = v_\varepsilon^{-1}\check{\tau}_{\varepsilon,v_\varepsilon,n}$, $n = 0, 1, \ldots$ and, also, that all conditions of Theorem 3.2 hold for the regenerative processes $(\xi_{\varepsilon,v_\varepsilon}(t), \eta_{\varepsilon,v_\varepsilon}(t)) = (\xi_\varepsilon(tv_\varepsilon), \eta_\varepsilon(tv_\varepsilon))$, $t \geq 0$ with the transition period $[0, \check{\tau}'_{\varepsilon,v_\varepsilon,1})$ and regeneration times $\check{\tau}'_{\varepsilon,v_\varepsilon,n} = v_\varepsilon^{-1}\check{\tau}'_{\varepsilon,v_\varepsilon,n}$, $n = 0, 1, \ldots$.

Note that, the regenerative lifetimes are not involved. The corresponding stopping probabilities $f_{\varepsilon,v_\varepsilon} = 0$, $\varepsilon \in (0, 1]$.

First, let us analyse the asymptotic behaviour of probabilities $\check{q}_{\varepsilon,v_\varepsilon,1}(t, A)$. Here, we can use the quasi-stationary ergodic relation given in Theorem 3.3.

Let us introduce random variables $\mu_{\varepsilon,1,n} = \kappa_{\varepsilon,1,n}\mathrm{I}(\eta_{\varepsilon,1,n} = 1)$, $n = 1, 2, \ldots$. Let now consider the sequence of random triplets $\langle\bar{\xi}_{\varepsilon,1,n} = \langle\xi_{\varepsilon,1,n}(t), t \geq 0\rangle, \kappa_{\varepsilon,1,n}, \mu_{\varepsilon,1,n}\rangle$, $n = 1, 2, \ldots$, the regenerative process $\xi_{\varepsilon,1}(t) = \xi_{\varepsilon,1,n}(t - \tau_{\varepsilon,1,n-1})$, for $t \in [\tau_{\varepsilon,1,n-1}, \tau_{\varepsilon,1,n})$, $n = 1, 2, \ldots$, with regeneration times $\tau_{\varepsilon,1,n} = \kappa_{\varepsilon,1,1} + \cdots + \kappa_{\varepsilon,1,n}$, $n = 1, 2, \ldots$, $\tau_{\varepsilon,1,0} = 0$, and the regenerative lifetime $\mu_{\varepsilon,1,+} = \tau_{\varepsilon,1,\nu_{\varepsilon,1}}$, where $\nu_{\varepsilon,1} = \min(n \geq 1 : \mu_{\varepsilon,1,n} < \kappa_{\varepsilon,1,n}) = \min(n \geq 1 : \eta_{\varepsilon,1,n} = 2)$.

Let us also denote $P_{\varepsilon,1,+}(t, A) = \mathsf{P}_1\{\xi_{\varepsilon,1}(t) \in A, \mu_{\varepsilon,1,+} > t\}$. In this case, the distribution functions $\bar{F}_{\varepsilon,1}(t) = \mathsf{P}\{\kappa_{\varepsilon,1,1} \leq t\}$ and $F_{\varepsilon,1}(t) = \mathsf{P}\{\kappa_{\varepsilon,1,1} \leq t, \mu_{\varepsilon,1,1} \geq \kappa_{\varepsilon,1,1}\} = \mathsf{P}\{\kappa_{\varepsilon,1,1} \leq t, \eta_{\varepsilon,1,1} = 1\}$, the stopping probability $f_{\varepsilon,1} = \mathsf{P}\{\mu_{\varepsilon,1,1} < \kappa_{\varepsilon,1,1}\} = \mathsf{P}\{\eta_{\varepsilon,1,1} = 2\} = p_{\varepsilon,12}$, and the expectations $\bar{e}_{\varepsilon,1} = \mathsf{E}\kappa_{\varepsilon,1,1} = e_{\varepsilon,11} + e_{\varepsilon,12}$ and $e_{\varepsilon,1} = \mathsf{E}\kappa_{\varepsilon,1,1}\mathrm{I}(\mu_{\varepsilon,1,1} \geq \kappa_{\varepsilon,1,1}) = \mathsf{E}\kappa_{\varepsilon,1,1}\mathrm{I}(\eta_{\varepsilon,1,1} = 1) = e_{\varepsilon,11}$.

It is also readily seen that, for every $A \in \mathcal{B}_{\mathbb{X}}$, $t \geq 0$,

$$\begin{aligned}\check{q}_{\varepsilon,1}(t, A) &= \mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \check{\tau}_{\varepsilon,1} > t\} \\ &= \mathsf{P}_1\{\xi_{\varepsilon,1}(t) \in A, \mu_{\varepsilon,1,+} > t\} = P_{\varepsilon,1,+}(t, A). \end{aligned} \tag{3.80}$$

Conditions **F**–**J** and $\mathbf{N}_1$ imply that conditions **A**–**D** holds for the regenerative processes $\xi_{\varepsilon,1}(t)$, $t \geq 0$ with regenerative times $\tau_{\varepsilon,1,n}$, $n = 1, 2, \ldots$ and regenerative lifetimes $\mu_{\varepsilon,1,+}$.

Let $s \in (0, \infty)$. We choose an arbitrary $0 \leq s_\varepsilon \to s$ as $\varepsilon \to 0$.

The above relation obviously implies that $s_\varepsilon v_\varepsilon \to \infty$. Conditions $\mathbf{N}_1$ and $\mathbf{K}_\beta$ obviously imply that $f_{\varepsilon,1}s_\varepsilon v_\varepsilon = p_{\varepsilon,12}s_\varepsilon v_\varepsilon = s_\varepsilon(1 + p_{\varepsilon,12}/p_{\varepsilon,21}) \to t_{s,\beta} = s(1 + \beta)$ as $\varepsilon \to 0$. Note that $t_{s,\beta} \in (0, \infty)$, for $\beta \in [0, \infty)$, while $t_{s,\infty} = \infty$.

Thus, all conditions of Theorem 3.3 hold for the regenerative processes $\xi_{\varepsilon,1}(t)$, $t \geq 0$ with the regenerative times $\tau_{\varepsilon,1,n}$, $n = 1, 2, \ldots$ and the regenerative lifetimes $\mu_{\varepsilon,1,+}$. Therefore, the following relation holds, for any $A \in \Gamma$, and $s \in (0, \infty)$,

$$\begin{aligned}P_{\varepsilon,1,+}(s_\varepsilon v_\varepsilon, A) &= \check{q}_{\varepsilon,1}(s_\varepsilon v_\varepsilon, A) = \check{q}_{\varepsilon,v_\varepsilon,1}(s_\varepsilon, A) \\ &\to \check{q}_{0,v_0,1}(s, A) = e^{-t_{s,\beta}/e_{0,1}}\pi_{0,1}(A) \text{ as } \varepsilon \to 0. \end{aligned} \tag{3.81}$$

If $\beta \in [0, \infty)$, then the limiting function $\check{q}_{0,v_0,1}(s, A) = e^{-s(1+\beta)/e_{0,1}}\pi_{0,1}(A)$, $s \in (0, \infty)$ is a non-trivial exponential function. However, if $\beta = \infty$, the limiting function $\check{q}_{0,v_0,1}(s, A) = 0$, $s \in (0, \infty)$.

In both cases, we can define $\check{q}_{0,v_0,1}(0, A) = \lim_{0<s\to 0}\check{q}_{0,v_0,1}(s, A)$.

Obviously, $\check{q}_{0,v_0,1}(0, A) = \pi_{0,1}(A)$, if $\beta \in [0, \infty)$, and $\check{q}_{0,v_0,1}(0, A) = 0$, if $\beta = \infty$.

Recall the limiting distribution function $\breve{Q}_{0,v_0,11}(t) = \mathsf{P}\{e_{0,1}\frac{1}{1+\beta}\zeta_1 + e_{0,2}\frac{1}{1+\beta^{-1}}\zeta_2 \le t\}$, $t \ge 0$ given by relation (3.66). Here, ζ_1 and ζ_2 are two independent random variables, exponentially distributed, with parameter 1.

Now, we are prepared to define the corresponding limiting regenerative process $(\xi_{0,v_0}(t), \eta_{0,v_0}(t))$, $t \ge 0$ with regeneration times $\breve{\tau}_{0,v_0,n}$, $n = 0, 1, \ldots$.

Let us $\xi_{i,n}$, $\kappa_{i,n}$, $i = 1, 2$, $n = 1, 2, \ldots$ be random variables, for which we assume that: (a) they are mutually independent; (b) their distributions do not depend on $n \ge 1$; (c) $\xi_{i,n}$ are random variables taking values in space $\mathbb{X}$ and such that, $\mathsf{P}\{\xi_{i,1} \in A\} = \pi_i(A)$, $A \in \mathcal{B}_{\mathbb{X}}$, for $i = 1, 2$; (d) the random variables $\kappa_{i,n}$ are non-negative random variables and $\mathsf{P}\{\kappa_{1,1} \le t\} = \mathsf{P}\{e_{0,1}\frac{1}{1+\beta}\zeta_1 \le t\}$, $t \ge 0$ while $\mathsf{P}\{\kappa_{2,1} \le t\} = \mathsf{P}\{e_{0,2}\frac{1}{1+\beta^{-1}}\zeta_2 \le t\}$, $t \ge 0$.

Now, let us define the inter-regeneration times $\kappa_{0,v_0,n} = \kappa_{1,n} + \kappa_{2,n}$, $n = 1, 2, \ldots$, the regeneration times $\tau_{0,v_0,n} = \kappa_{0,v_0,1} + \cdots + \kappa_{0,v_0,n}$, $n = 1, 2, \ldots$, $\tau_{0,v_0,0} = 0$, and the regeneration process $\xi_{0,v_0}(t) = \xi_{1,n}$, $\eta_{0,v_0}(t) = 1$, for $t \in [\tau_{0,v_0,n-1}, \tau_{0,v_0,n-1} + \kappa_{1,n})$ and $\xi_{0,v_0}(t) = \xi_{2,n}$, $\eta_{0,v_0}(t) = 2$, for $t \in [\tau_{0,v_0,n-1} + \kappa_{1,n}, \tau_{0,v_0,n})$, for $n = 0, 1, \ldots$.

It is readily seen that $\mathsf{P}\{\xi_{0,v_0}(t) \in A, \tau_{0,v_0,1} > t\} = \breve{q}_{0,v_0,1}(t, A)$, $t \ge 0$ and $\mathsf{P}\{\tau_{0,v_0,1} \le t\} = \breve{Q}_{0,v_0,11}(t)$, $t \ge 0$, where $\breve{q}_{0,v_0,1}(t, A)$ and $\breve{Q}_{0,v_0,11}(t)$ are given, respectively, by relations (3.81) and (3.66).

Therefore, the renewal equation (3.7) for probabilities $P_{0,v_0,11}(t, A) = \mathsf{P}\{\xi_{0,v_0}(t) \in A, \eta_{0,v_0}(t) = 1\}$ takes the following form,

$$P_{0,v_0,11}(t, A) = \breve{q}_{0,v_0,1}(t, A) + \int_0^t P_{0,v_0,11}(t - s, A)\breve{Q}_{0,v_0,11}(t), \ t \ge 0. \tag{3.82}$$

Conditions of Theorem 3.1 are satisfied for the regenerative process

$$(\xi_{\varepsilon,v_\varepsilon}(t), \eta_{\varepsilon,v_\varepsilon}(t)), t \ge 0$$

with regeneration times $\tau_{\varepsilon,v_\varepsilon,n}$, $n = 0, 1, \ldots$.

Indeed, condition **F** implies that condition **A** holds for the above regenerative processes. Relation (3.66) and conditions **G**, **H**, **I** and $\mathbf{N}_1$ imply that condition **B** holds. Relation (3.69) and conditions and conditions **G**, **H**, **I** and $\mathbf{N}_1$ also imply that condition **C** holds.

Due to an arbitrary choice of $0 \le s_\varepsilon \to s$ as $\varepsilon \to 0$, convergence in relation (3.81) is locally uniform in every point $s \in (0, \infty)$. Thus, by Lemma 3.1 given Sect. 3.4.3, the asymptotic relation in condition **D** holds for functions $\breve{q}_{\varepsilon,v_\varepsilon,1}(s, A)$, $s \in [0, \infty)$ for any $s \in (0, \infty)$.

Convergence at point 0 is not guarantied. However, $m(\{0\}) = 0$. Thus, condition **D** holds for functions $\breve{q}_{\varepsilon,v_\varepsilon,1}(s_\varepsilon, A)$, $s \in [0, \infty)$, with the limiting function $\breve{q}_{0,v_0,1}(s, A) = e^{-t_{s,\beta}/e_{0,11}}\pi_{0,1}(A)$, $s \in [0, \infty)$.

By the above remarks, all conditions of Theorem 3.1 hold, and the ergodic relation given in this theorem takes place for probabilities $P_{\varepsilon,v_\varepsilon,11}(t'_\varepsilon, A) = P_{\varepsilon,11}(t'_\varepsilon v_\varepsilon, A)$ for any $0 \le t'_\varepsilon \to \infty$ as $\varepsilon \to 0$,

$$P_{\varepsilon,v_\varepsilon,11}(t'_\varepsilon, A) = P_{\varepsilon,11}(t'_\varepsilon v_\varepsilon, A) \to \pi^{(\beta)}_{0,v_0,1}(A)$$
$$= \frac{1}{\breve{e}_{0,v_0,11}} \int_0^\infty e^{-t_{s,\beta}/e_{0,1}} \pi_{0,1}(A) m(ds) \text{ as } \varepsilon \to 0. \tag{3.83}$$

Relation (3.69) and formula $t_{s,\beta} = s(1+\beta)$ imply that probabilities $\pi^{(\beta)}_{0,v_0,1}(A)$ coincide with probabilities $\pi^{(\beta)}_{0,1}(A)$ given in relation (3.78). Indeed,

$$\pi^{(\beta)}_{0,v_0,1}(A) = \frac{1}{\breve{e}_{0,v_0,11}} \int_0^\infty e^{-s(1+\beta)/e_{0,1}} \pi_{0,1}(A) m(ds)$$
$$= \frac{e_{0,1}(1+\beta)^{-1}}{e_{0,1}(1+\beta)^{-1} + e_{0,2}(1+\beta^{-1})^{-1}} \pi_{0,1}(A) = \pi^{(\beta)}_{0,1}(A). \tag{3.84}$$

Thus, the following ergodic relation holds for any for $A \in \Gamma$ and $0 \le t'_\varepsilon \to \infty$ as $\varepsilon \to 0$,

$$P_{\varepsilon,11}(t'_\varepsilon v_\varepsilon, A) \to \pi^{(\beta)}_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.85}$$

Let us now consider the compressed version of the regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t)), t \ge 0$ with the transition period $[0, \breve{\tau}'_{\varepsilon,1})$ with regeneration times $\breve{\tau}'_{\varepsilon,n}, n = 0, 1, \ldots$. It is the regenerative process $(\xi_{\varepsilon,v_\varepsilon}(t), \eta_{\varepsilon,v_\varepsilon}(t)), t \ge 0 = (\xi_\varepsilon(tv_\varepsilon), \eta_\varepsilon(tv_\varepsilon)), t \ge 0$ with regeneration times $\tau'_{\varepsilon,v_\varepsilon,n} = v_\varepsilon^{-1}\breve{\tau}'_{\varepsilon,n}, n = 0, 1, \ldots$.

The shifted process $(\xi_{\varepsilon,v_\varepsilon}(\breve{\tau}'_{\varepsilon,v_\varepsilon,1} + t), \eta_{\varepsilon,v_\varepsilon}(\breve{\tau}'_{\varepsilon,v_\varepsilon,1} + t)), t \ge 0$ is a standard regenerative process. If $\eta_{\varepsilon,v_\varepsilon}(0) = 2$, then $\eta_{\varepsilon,v_\varepsilon}(\tilde{\tau}_{\varepsilon,v_\varepsilon,1}) = 1$. That is why, probabilities $P_{\varepsilon,v_\varepsilon,11}(t, A)$ play for this process the role of probabilities $P^{(1)}_\varepsilon(t, A)$ pointed out in Sect. 3.2.1.

Relation (3.65) and conditions **G**, **H**, **I** and $\mathbf{N}_1$ imply that condition **E** holds for the distribution functions $\tilde{Q}_{\varepsilon,v_\varepsilon,21}(t) = \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,v_\varepsilon,1} \le t\}, t \ge 0$.

Thus, all conditions of Theorem 3.2 hold, and the ergodic relation (3.85) for probabilities $P_{\varepsilon,v_\varepsilon,11}(t'_\varepsilon, A) = P_{\varepsilon,11}(t'_\varepsilon v_\varepsilon, A)$ also holds for probabilities $P_{\varepsilon,v_\varepsilon,21}(t'_\varepsilon, A) = P_{\varepsilon,21}(t'_\varepsilon v_\varepsilon, A)$.

Due to the symmetricity of conditions **G**–**J**, $\mathbf{K}_\beta$, and $\mathbf{N}_1$ with respect to the indices $i, j = 1, 2$, the ergodic relations, analogous to the mentioned above ergodic relations for probabilities $P_{\varepsilon,v_\varepsilon,11}(t'_\varepsilon, A) = P_{\varepsilon,11}(t'_\varepsilon v_\varepsilon, A)$, $P_{\varepsilon,v_\varepsilon,21}(t'_\varepsilon, A) = P_{\varepsilon,21}(t'_\varepsilon v_\varepsilon, A)$ also hold for probabilities

$$P_{\varepsilon,v_\varepsilon,22}(t'_\varepsilon, A) = P_{\varepsilon,22}(t'_\varepsilon v_\varepsilon, A), \quad P_{\varepsilon,v_\varepsilon,12}(t'_\varepsilon, A) = P_{\varepsilon,12}(t'_\varepsilon v_\varepsilon, A).$$

They have the following forms, $P_{\varepsilon,v_\varepsilon,i2}(t'_\varepsilon, A) = P_{\varepsilon,i2}(t'_\varepsilon v_\varepsilon, A) \to \pi^{(\beta)}_{0,2}(A)$ as $\varepsilon \to 0$, for $i = 1, 2$.

The above analysis, in particular, relation (3.85), yields the description of asymptotic behaviour of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ for super-long times $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ satisfying the asymptotic relation $t_\varepsilon / v_\varepsilon \to \infty$ as $\varepsilon \to 0$. To see this, one should just represent such t_ε in the form, $t_\varepsilon = t'_\varepsilon v_\varepsilon$. Obviously, $t'_\varepsilon = t_\varepsilon / v_\varepsilon \to \infty$ as $\varepsilon \to 0$. □

3.4.5 Long Time Ergodic Theorems for Singularly Perturbed Alternating Regenerative Processes

In this subsection, we describe the asymptotic behaviour of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ for so-called "long" times $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$, which satisfy asymptotic relation,

$$t_\varepsilon / v_\varepsilon \to t \in (0, \infty) \text{ as } \varepsilon \to 0. \tag{3.86}$$

Let $\beta \in (0, \infty)$, and $\eta^{(\beta)}(t), t \ge 0$ be a homogeneous continuous time Markov chain with the phase space $\mathbb{Y} = \{1, 2\}$, the transition probabilities of embedded Markov chain $p_{ij} = \mathrm{I}(i \ne j), i, j = 1, 2$, and the distribution functions of sojourn times in states 1 and 2, respectively, $F_1^{(\beta)}(t) = 1 - e^{-t(1+\beta)/e_{0,1}}, t \ge 0$ and $F_2^{(\beta)}(t) = 1 - e^{-t(1+\beta^{-1})/e_{0,2}}, t \ge 0$. We also assume that this Markov chain has continuous from the right trajectories.

Let us $p_{ij}^{(\beta)}(t) = \mathsf{P}_i\{\eta^{(\beta)}(t) = j\}, t \ge 0, i, j = 1, 2$ be transition probabilities for the Markov chain $\eta^{(\beta)}(t)$.

The explicit expression for the transition probabilities $p_{ij}^{(\beta)}(t)$ are well known, as the solutions of the corresponding forward Kolmogorov system of differential equations for these probabilities. Namely, the corresponding matrix $\| p_{ij}^{(\beta)}(t) \|$ has the following form, for $t \ge 0$,

$$\| p_{ij}^{(\beta)}(t) \| = \left\| \begin{array}{ll} \rho_1(\beta) + \rho_2(\beta) e^{-\lambda(\beta)t} & \rho_2(\beta) - \rho_2(\beta) e^{-\lambda(\beta)t} \\ \rho_1(\beta) - \rho_1(\beta) e^{-\lambda(\beta)t} & \rho_2(\beta) + \rho_1(\beta) e^{-\lambda(\beta)t} \end{array} \right\|, \tag{3.87}$$

where

$$\lambda_1(\beta) = \frac{1+\beta}{e_{0,1}}, \ \lambda_2(\beta) = \frac{1+\beta^{-1}}{e_{0,2}}, \ \lambda(\beta) = \lambda_1(\beta) + \lambda_2(\beta), \tag{3.88}$$

and

$$\rho_1(\beta) = \frac{\lambda_2(\beta)}{\lambda(\beta)} = \frac{e_{0,1}(1+\beta)^{-1}}{e(\beta)}, \ \rho_2(\beta) = \frac{\lambda_1(\beta)}{\lambda(\beta)} = \frac{e_{0,2}(1+\beta^{-1})^{-1}}{e(\beta)}. \tag{3.89}$$

Note that the Markov chain $\eta^{(\beta)}(t)$ is ergodic and $\rho_i(\beta), i = 1, 2$ are its stationary probabilities.

The corresponding limiting probabilities have in this case the following forms, for $A \in \mathcal{B}_{\mathbb{X}}, \ i, j = 1, 2, \ t \in (0, \infty)$,

$$\pi_{0,i1}^{(\beta)}(t, A) = \left\{ \begin{array}{cc} \pi_{0,1}(A) & \text{for } \ i = 1, 2, \beta = 0, \\ p_{i1}^{(\beta)}(t) \pi_{0,1}(A) & \text{for } \ i = 1, 2, \beta \in (0, \infty), \\ 0 & \text{for } \ i = 1, 2, \beta = \infty, \end{array} \right. \tag{3.90}$$

and

$$\pi_{0,i2}^{(\beta)}(t, A) = \begin{cases} 0 & \text{for } i = 1, 2, \beta = 0, \\ p_{i2}^{(\beta)}(t)\pi_{0,2}(A) & \text{for } i = 1, 2, \beta \in (0, \infty), \\ \pi_{0,2}(A) & \text{for } i = 1, 2, \beta = \infty. \end{cases} \tag{3.91}$$

The following theorem takes place.

Theorem 3.9 *Let conditions* **F**–**J**, $\mathbf{N}_1$ *hold and, also, condition* $\mathbf{K}_\beta$ *holds for some* $\beta \in [0, \infty]$. *Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / v_\varepsilon \to t \in (0, \infty)$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,ij}^{(\beta)}(t, A) \text{ as } \varepsilon \to 0. \tag{3.92}$$

Proof Let us again us consider the renewal equation (3.82) for the compressed regenerative process $(\xi_{\varepsilon,v_\varepsilon}(t), \eta_{\varepsilon,v_\varepsilon}(t)), t \ge 0 = (\xi_\varepsilon(tv_\varepsilon), \eta_\varepsilon(tv_\varepsilon)), t \ge 0$ with regeneration times $\check{\tau}_{\varepsilon,v_\varepsilon,n} = v_\varepsilon^{-1}\check{\tau}_{\varepsilon,n}, n = 0, 1, \ldots$.

As well known, the solution of this equation has the form,

$$P_{\varepsilon,v_\varepsilon,11}(t, A) = \int_0^t \check{q}_{\varepsilon,v_\varepsilon,1}(t - s, A)\check{U}_{\varepsilon,v_\varepsilon,11}(ds), t \ge 0, \tag{3.93}$$

where

$$\check{U}_{\varepsilon,v_\varepsilon,11}(t) = \sum_{n=0}^{\infty} \check{Q}_{\varepsilon,v_\varepsilon,11}^{*n}(t), \ t \ge 0, \tag{3.94}$$

is the corresponding renewal function.

Inequality $\check{Q}_{\varepsilon,v_\varepsilon,11}^{*n}(t) \le \check{Q}_{\varepsilon,v_\varepsilon,11}^{n}(t)$ obviously holds for any $t \ge 0$ and $n \ge 1$. These inequalities and relation (3.66) imply that, $\overline{\lim}_{\varepsilon\to 0}\, \check{Q}_{\varepsilon,v_\varepsilon,11}^{*n}(t) \le \overline{\lim}_{\varepsilon\to 0}\, \check{Q}_{\varepsilon,v_\varepsilon,11}^{n}(t) = \tilde{Q}_{0,v_0,11}^{n}(t) < 1$, since $\tilde{Q}_{0,v_0,11}(t) = \mathsf{P}\{e_{0,1}\frac{1}{1+\beta}\zeta_1 + e_{0,2}\frac{1}{1+\beta^{-1}}\zeta_2 \le t\} < 1$. Thus, the series on the right hand side in (3.94) converge asymptotically uniformly, as $\varepsilon \to 0$.

Also, relation (3.66) implies that $\check{Q}_{\varepsilon,v_\varepsilon,11}^{*n}(\cdot) \Rightarrow \check{Q}_{0,v_0,11}^{*n}(\cdot)$ as $\varepsilon \to 0$.

The above remarks imply that, for $t > 0$,

$$\check{U}_{\varepsilon,v_\varepsilon,11}(t) \to \check{U}_{0,v_0,11}(t) \text{ as } \varepsilon \to 0. \tag{3.95}$$

The convergence relation in (3.95) holds for all $t > 0$, since $\tilde{Q}_{0,v_0,11}^{*n}(t), t \ge 0$ is a continuous distribution function and, in sequel, due to the above remarks, $\check{U}_{0,v_0,11}(t), t \ge 0$ is continuous function.

Relation (3.81) implies that, for every $t > 0$, functions $\check{q}_{\varepsilon,v_\varepsilon,1}(t - s, A) \xrightarrow{\mathrm{U}} \check{q}_{0,v_0,1}(t - s, A)$ as $\varepsilon \to 0$, for $s \in [0, t)$. At the same time, due to continuity function $\check{U}_{0,v_0,11}(t)$, for $t > 0$, measure $\check{U}_{0,v_0,11}(ds)$ has no atom at any point $t > 0$.

By the above remarks and relations (3.81), (3.95), Lemma 3.2 formulated in Sect. 3.4.3 imply, that the following relation holds, for $A \in \Gamma$ and $t > 0$,

$$P_{\varepsilon,v_\varepsilon,11}(t,A)=\int_0^t \check{q}_{\varepsilon,v_\varepsilon,1}(t-s,A)\check{U}_{\varepsilon,v_\varepsilon,11}(ds)$$
$$\to P_{0,v_0,11}(t,A)=\int_0^t \check{q}_{0,v_0,1}(t-s,A)\check{U}_{0,v_0,11}(ds)$$
$$=\pi_1(A)\int_0^t e^{-(t-s)(1+\beta)/e_{0,1}}\check{U}_{0,v_0,11}(ds) \text{ as } \varepsilon\to 0. \tag{3.96}$$

Next, we make an important remark that the scaling of time transformation with the compression factors v_ε and all following asymptotic relations presented above can be, in obvious way, repeated for any slightly modified compression factors $\dot{v}_\varepsilon = a_\varepsilon v_\varepsilon$, where $0 < a_\varepsilon \to 1$ as $\varepsilon \to 0$.

In particular, the modified asymptotic relation (3.96) takes the following form, for $A \in \Gamma$ and $t > 0$,

$$P_{\varepsilon,\dot{v}_\varepsilon,11}(t,A)=P_{\varepsilon,11}(ta_\varepsilon v_\varepsilon,A)\to P_{0,v_0,11}(t,A) \text{ as } \varepsilon\to 0. \tag{3.97}$$

Due to an arbitrary choice of $0 < a_\varepsilon \to 1$ as $\varepsilon \to 0$, relation (3.97) is, for every $t > 0$, equivalent to the following relation, which holds for any $A \in \Gamma$ and $0 \le t''_\varepsilon \to t$ as $\varepsilon \to 0$,

$$P_{\varepsilon,11}(t''_\varepsilon v_\varepsilon,A)=P_{\varepsilon,v_\varepsilon,11}(t''_\varepsilon,A)\to P_{0,v_0,11}(t,A) \text{ as } \varepsilon\to 0. \tag{3.98}$$

The shifted process $(\xi_{\varepsilon,v_\varepsilon}(\check{\tau}'_{\varepsilon,v_\varepsilon,1}+t),\eta_{\varepsilon,v_\varepsilon}(\check{\tau}'_{\varepsilon,v_\varepsilon,1}+t))$, $t \ge 0$ is a standard regenerative process. If $\eta_{\varepsilon,v_\varepsilon}(0)=2$, then $\eta_{\varepsilon,v_\varepsilon}(\tilde{\tau}_{\varepsilon,v_\varepsilon,1})=1$. That is why, probabilities $P_{\varepsilon,v_\varepsilon,11}(t,A)$ play for this process the role of probabilities $P^{(1)}_\varepsilon(t,A)$ pointed out in Sect. 3.2.1. The distribution function for the duration of transition period is $\tilde{Q}_{\varepsilon,v_\varepsilon,21}(t)=\mathsf{P}_2\{\tilde{\tau}_{\varepsilon,v_\varepsilon,1}\le t\}$, $t \ge 0$.

According relation (3.65), the distribution functions $\tilde{Q}_{\varepsilon,v_\varepsilon,21}(t)$ weakly converge as $\varepsilon \to 0$ to the distribution function $\tilde{Q}_{0,v_0,21}(t)=\mathsf{P}\{e_{0,2}\frac{1}{1+\beta^{-1}}\zeta\le t\}$ which is continuous function for $t > 0$. If $\beta \in (0,\infty]$, then $\tilde{Q}_{0,v_0,21}(t)=1-e^{-t(1+\beta^{-1})/e_{0,2}}$, $t \ge 0$ is the exponential distribution function. If $\beta = 0$, then $\tilde{Q}_{0,v_0,21}(t)=\mathrm{I}(t\ge 0)$, $t \ge 0$.

The renewal type transition relation (3.11) takes the following form,

$$P_{\varepsilon,v_\varepsilon,21}(t,A)=\int_0^t P_{\varepsilon,v_\varepsilon,11}(t-s,A)\tilde{Q}_{\varepsilon,v_\varepsilon,21}(ds),\ t\ge 0. \tag{3.99}$$

Relation (3.98) implies that, for every $t > 0$, functions $P_{\varepsilon,v_\varepsilon,11}(t-s,A) \xrightarrow{\mathrm{U}} P_{0,v_0,11}(t-s,A)$ as $\varepsilon \to 0$, for $s \in [0,t)$. At the same time, due to continuity the distribution function $\tilde{Q}_{0,v_0,21}(t)$ for $t > 0$ measure $\tilde{Q}_{0,v_0,21}(ds)$ has no atom at any point $t > 0$. By these remarks and relations (3.65), (3.98), Lemma 3.2 formulated in Sect. 3.4.3 implies that the following relation holds, for $A \in \Gamma$ and $t > 0$,

$$P_{\varepsilon,v_\varepsilon,21}(t,A) = \int_0^t P_{\varepsilon,v_\varepsilon,11}(t-s,A)\tilde{Q}_{\varepsilon,v_\varepsilon,21}(ds)$$
$$\to \int_0^t P_{0,v_0,11}(t-s,A)\tilde{Q}_{0,v_0,21}(ds) = P_{0,v_0,21}(t,A). \quad (3.100)$$

By arguments similar with those used for relations (3.96)–(3.98), one can, for every $t > 0$, improve relation (3.100) to the more advanced form of this relation, which holds for $A \in \Gamma$ and any $0 \le t''_\varepsilon \to t$ as $\varepsilon \to 0$,

$$P_{\varepsilon,21}(t''_\varepsilon v_\varepsilon, A) = P_{\varepsilon,v_\varepsilon,21}(t''_\varepsilon, A) \to P_{0,v_0,21}(t,A) \text{ as } \varepsilon \to 0. \quad (3.101)$$

It is remains to give a more explicit expression for the limiting probabilities $P_{0,v_0,11}(t,A)$, $t > 0$ and $P_{0,v_0,21}(t,A)$, $t > 0$.

First, let us consider the case, where $\beta = 0$.

In this case, $\check{Q}_{0,v_0,11}(t) = \mathsf{P}\{e_{0,1}\zeta_1 \le t\} = 1 - e^{-t/e_{0,1}}$, $t \ge 0$, $i = 1, 2$ is an exponential distribution function. Thus, the renewal function $\check{U}_{0,v_0,11}(t) = \mathrm{I}(t \ge 0) + \frac{1}{e_{0,1}}t$, $t \ge 0$. Also, $\tilde{Q}_{0,v_0,21}(t) = \mathrm{I}(t \ge 0)$, $t \ge 0$. Finally, $t_{s,0} = s$, $s \ge 0$. That is why, for $A \in \mathcal{B}_{\mathbb{X}}$ and $t > 0$,

$$P_{0,v_0,11}(t,A) = \pi_{0,1}(A)\int_0^t e^{-(t-s)/e_{0,1}}\check{U}_{0,v_0,11}(ds)$$
$$= \pi_{0,1}(A)(e^{-t/e_{0,1}} + \int_0^t \frac{e^{-(t-s)/e_{0,1}}}{e_{0,1}}ds)$$
$$= \pi_{0,1}(A)(e^{-t/e_{0,1}} + e^{-t/e_{0,1}}(e^{t/e_{0,1}} - 1)) = \pi_{0,1}(A). \quad (3.102)$$

and

$$P_{0,v_0,21}(t,A) = \int_0^t P_{0,v_0,11}(t-s,A)\tilde{Q}_{0,v_0,21}(ds) = \pi_{0,1}(A). \quad (3.103)$$

Second, let us consider the case, where $\beta = \infty$.

In this case, $\tilde{Q}_{0,v_0,11}(t) = \mathsf{P}\{e_{0,2}\zeta_2 \le t\} = 1 - e^{-t/e_{0,2}}$, $t \ge 0$ is an exponential distribution function. Thus, the renewal function $\check{U}_{0,v_0,11}(t) = \mathrm{I}(t \ge 0) + \frac{1}{e_{0,2}}t$, $t \ge 0$. Also, $\tilde{Q}_{0,v_0,21}(t) = \mathsf{P}\{e_{0,2}\zeta_2 \le t\} = 1 - e^{-t/e_{0,2}}$, $t \ge 0$. Finally, $t_{s,\infty} = \infty$, $s \ge 0$. That is why, for $t > 0$,

$$P_{0,v_0,11}(t,A) = P_{0,v_0,21}(t,A) = 0. \quad (3.104)$$

Third, let us consider the main case, where $\beta \in (0,\infty)$.

Let us $\eta^{(\beta)}(t)$, $t \ge 0$ be a continuous time homogeneous Markov chain introduced in the beginning of this subsection.

Let $\tau_n^{(\beta)} = \inf(t > \tau_{n-1}^{(\beta)}, \eta^{(\beta)}(t) \ne \eta^{(\beta)}(\tau_{n-1}^{(\beta)}))$, $n = 1, 2, \ldots$, $\tau_0^{(\beta)} = 0$ be the sequential moments of jumps for the Markov chain $\eta^{(\beta)}(t)$.

The Markov chain $\eta^{(\beta)}(t)$ obviously is also an alternating regenerative process, with regeneration times $\tau_{2n}^{(\beta)}$, $n = 0, 1, \ldots$.

Let us assume that $\eta^{(\beta)}(0) = 1$. The transition probabilities $p_{11}^{(\beta)}(t)$, $t \geq 0$ satisfy the following renewal equation,

$$p_{11}^{(\beta)}(t) = q_1^{(\beta)}(t) + \int_0^t p_{11}^{(\beta)}(t-s)F_{11}^{(\beta)}(ds), \ t \geq 0, \tag{3.105}$$

where $q_1^{(\beta)}(t) = \mathsf{P}_1\{\eta^{(\beta)}(t) = 1, \tau_2^{(\beta)} > t\}$ and $F_{11}^{(\beta)}(t) = \mathsf{P}_1\{\tau_2^{(\beta)} \leq t\}$, fior $t \geq 0$.

Let $U_{11}^{(\beta)}(t) = \sum_{n=0}^{\infty} F_{11}^{(\beta)*n}(t)$, $t \geq 0$ be the corresponding renewal function generated by the distribution function $F_{11}^{(\beta)}(t)$. The transition probabilities $p_{11}^{(\beta)}(t)$ can be expressed as the solution of the renewal equation (3.105) in the following form,

$$p_{11}^{(\beta)}(t) = \int_0^t q_1^{(\beta)}(t)U_{11}^{(\beta)}(ds), \ t \geq 0. \tag{3.106}$$

Obviously,

$$q_1^{(\beta)}(t) = \mathsf{P}_1\{\tau_1^{(\beta)} > t\} = e^{-t(1+\beta)/e_{0,1}}, \ t \geq 0, \tag{3.107}$$

and

$$F_{11}^{(\beta)}(t) = F_1^{(\beta)}(t) * F_2^{(\beta)}(t) = \check{Q}_{0,v_0,11}(t), \ t \geq 0. \tag{3.108}$$

and, thus,

$$U_{11}^{(\beta)}(t) = \check{U}_{0,v_0,11}(t), \ t \geq 0. \tag{3.109}$$

Relations (3.106), (3.107), and (3.109) imply that,

$$p_{11}^{(\beta)}(t) = \int_0^t e^{-(t-s)(1+\beta)/e_{0,1}}\check{U}_{0,v_0,11}(ds), \ t \geq 0. \tag{3.110}$$

Finally, relations (3.96) and (3.110) imply that that the following equality takes place, for $t > 0$,

$$P_{0,v_0,11}(t, A) = p_{11}^{(\beta)}(t)\pi_{0,1}(A). \tag{3.111}$$

The distribution function $F_2^{(\beta)}(t) = \check{Q}_{0,v_0,21}(t) = 1 - e^{-t(1+\beta^{-1})/e_{0,2}}$, $t \geq 0$. Thus,

$$p_{21}^{(\beta)}(t) = \int_0^t p_{11}^{(\beta)}(t-s)\check{Q}_{0,v_0,21}(ds), t \geq 0, \tag{3.112}$$

and, therefore, relation (3.100) implies that following equality takes place, for $t > 0$,

$$P_{0,v_0,21}(t, A) = p_{21}^{(\beta)}(t)\pi_{0,1}(A). \tag{3.113}$$

Due to the symmetricity of conditions **F**–**J**, $\mathbf{K}_\beta$, and $\mathbf{N}_1$ with respect to the indices $i, j = 1, 2$, the ergodic relations, analogous to the mentioned above ergodic relations for probabilities $P_{\varepsilon,v_\varepsilon,11}(t''_\varepsilon, A) = P_{\varepsilon,11}(t''_\varepsilon v_\varepsilon, A)$ and $P_{\varepsilon,v_\varepsilon,21}(t''_\varepsilon, A) = P_{\varepsilon,21}(t''_\varepsilon v_\varepsilon, A)$, also hold for probabilities

$$P_{\varepsilon,v_\varepsilon,22}(t''_\varepsilon, A) = P_{\varepsilon,22}(t''_\varepsilon v_\varepsilon, A), \;\; P_{\varepsilon,v_\varepsilon,12}(t''_\varepsilon, A) = P_{\varepsilon,12}(t''_\varepsilon v_\varepsilon, A).$$

The above analysis yields the description of asymptotic behaviour of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ for long times $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ satisfying the asymptotic relation $t_\varepsilon / v_\varepsilon \to t \in (0, \infty)$ as $\varepsilon \to 0$. To see this, one should just represent such t_ε in the form, $t_\varepsilon = t''_\varepsilon v_\varepsilon$. Obviously, $t''_\varepsilon = t_\varepsilon / v_\varepsilon \to t$ as $\varepsilon \to 0$. □

3.5 Short Time Ergodic Theorems for Singularly Perturbed Alternating Regenerative Processes

In this section, we present short time individual ergodic theorems for singularly perturbed alternating regenerative processes.

3.5.1 Short Time Ergodic Theorems for Singularly Perturbed Alternating Regenerative Processes - I

In this subsection, we describe the asymptotic behaviour of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ for so-called "short" times $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$, which satisfy the following asymptotic relation,

$$t_\varepsilon / v_\varepsilon \to 0 \text{ as } \varepsilon \to 0. \tag{3.114}$$

We also assume that, additionally to conditions $\mathbf{N}_1$, condition $\mathbf{K}_\beta$ holds for some $\beta \in (0, \infty)$.

The corresponding limiting probabilities are, in this case, the same for any $\beta \in (0, \infty)$ and take the following form, for $A \in \mathcal{B}_{\mathbb{X}}$, $i, j = 1, 2$,

$$\pi_{0,ij}(A) = \mathrm{I}(j = i)\pi_{0,i}(A) = \begin{cases} \pi_{0,i}(A) & \text{for } j = i, \\ 0 & \text{for } j \ne i. \end{cases} \tag{3.115}$$

The following theorem takes place.

Theorem 3.10 *Let conditions* **F**–**J**, $\mathbf{N}_1$ *hold and, also, condition* $\mathbf{K}_\beta$ *holds for some* $\beta \in (0, \infty)$. *Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / v_\varepsilon \to 0$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,ij}(A) \text{ as } \varepsilon \to 0. \tag{3.116}$$

Proof Let us start by analysing the asymptotic behaviour of probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ and, thus, assume that $\eta_\varepsilon(0) = 1$.

We return back to the initial alternating regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \geq 0$ with regeneration times $\tau_{\varepsilon,n}$, $n = 0, 1, \ldots$.

Recall the stopping time $\tilde{\tau}_{\varepsilon,1}$, which is the time of first hitting sate 2 by process $\eta_\varepsilon(t)$.

Let us again consider the regenerative process $\xi_{\varepsilon,1}(t)$, $t \geq 0$ with regeneration times $\tau_{\varepsilon,1,n}$, $n = 0, 1, \ldots$, and the random lifetime $\mu_{\varepsilon,1,+}$ introduced in Sect. 3.4.4.

It is readily seen that, for every $t \geq 0$,

$$\tilde{Q}_{\varepsilon,12}(t) = \mathsf{P}_1\{\tilde{\tau}_{\varepsilon,1} \leq t\} = \mathsf{P}\{\mu_{\varepsilon,1,+} \leq t\} \tag{3.117}$$

and, for every $A \in \mathcal{B}_{\mathbb{X}}$, $t \geq 0$,

$$\mathsf{P}_1\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 1, \tilde{\tau}_{\varepsilon,1} > t\} = \mathsf{P}\{\xi_{\varepsilon,1}(t) \in A, \mu_{\varepsilon,1,+} > t\}. \tag{3.118}$$

According relation (3.63), if $\eta_\varepsilon(0) = 1$, random variables,

$$v_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \xrightarrow{d} e_{0,1}\frac{1}{1+\beta}\zeta \text{ as } \varepsilon \to 0, \tag{3.119}$$

where ζ is a random variable exponentially distributed, with parameter 1.

Since, we assumed that $t_\varepsilon / v_\varepsilon \to 0$ as $\varepsilon \to 0$, relations (3.117) and (3.119) imply that,

$$\begin{aligned}\mathsf{P}\{\mu_{\varepsilon,1,+} > t_\varepsilon\} &= \mathsf{P}_1\{\tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} \\ &= \mathsf{P}_1\{v_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} > t_\varepsilon v_\varepsilon^{-1}\} \to 1 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.120}$$

Relations (3.118) and (3.120) imply that

$$\begin{aligned}&\mathsf{P}_1\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 1\} - \mathsf{P}_1\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 1, \tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} \\ &\quad \leq \mathsf{P}_1\{\tilde{\tau}_{\varepsilon,1} \leq t_\varepsilon\} \to 0 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.121}$$

and, analogously,

$$\begin{aligned}&\mathsf{P}\{\xi_{\varepsilon,1}(t_\varepsilon) \in A\} - \mathsf{P}\{\xi_{\varepsilon,1}(t_\varepsilon) \in A, \mu_{\varepsilon,1,+} > t_\varepsilon\} \\ &\quad \leq \mathsf{P}\{\mu_{\varepsilon,1,+} \leq t_\varepsilon\} \to 0 \text{ as } \varepsilon \to 0,\end{aligned} \tag{3.122}$$

These relations and Theorem 3.1, which can be applied to the regenerative processes $\xi_{\varepsilon,1}(t)$, imply that, for every $A \in \Gamma$,

$$\begin{aligned}\lim_{\varepsilon\to 0}\mathsf{P}_{11}(t_\varepsilon, A) &= \lim_{\varepsilon\to 0}\mathsf{P}_1\{\xi_\varepsilon(t_\varepsilon)\in A, \eta_\varepsilon(t_\varepsilon)=1\}\\ &= \lim_{\varepsilon\to 0}\mathsf{P}_1\{\xi_\varepsilon(t_\varepsilon)\in A, \eta_\varepsilon(t_\varepsilon)=1, \tilde{\tau}_{\varepsilon,1}>t_\varepsilon\}\\ &= \lim_{\varepsilon\to 0}\mathsf{P}\{\xi_{\varepsilon,1}(t_\varepsilon)\in A, \mu_{\varepsilon,1,+}>t_\varepsilon\}\\ &= \lim_{\varepsilon\to 0}\mathsf{P}\{\xi_{\varepsilon,1}(t_\varepsilon)\in A\} = \pi_{0,1}(A). \end{aligned} \tag{3.123}$$

Let us now analyse the asymptotic behaviour for probabilities $P_{\varepsilon,21}(t, A)$ and, thus, assume that $\eta_\varepsilon(0) = 2$.

In this case, relation (3.65) implies that random variables,

$$v_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \xrightarrow{d} e_{0,2}\frac{1}{1+\beta^{-1}}\zeta \text{ as } \varepsilon\to 0, \tag{3.124}$$

where ζ is a random variable exponentially distributed, with parameter 1.

Since, we assumed that $t_\varepsilon/v_\varepsilon \to 0$ as $\varepsilon\to 0$, the above convergence in distribution relation, obviously, implies that,

$$\mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1}>t_\varepsilon\} = \mathsf{P}_2\{v_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1}>t_\varepsilon v_\varepsilon^{-1}\}\to 1 \text{ as } \varepsilon\to 0. \tag{3.125}$$

If $\eta_\varepsilon(0) = 2$, then, for every $t > 0$, event $\{\eta_\varepsilon(t) = 1\} \subseteq \{\tilde{\tau}_{\varepsilon,1}\le t\}$. Thus, for every $A\in\Gamma$,

$$\begin{aligned}\mathsf{P}_{21}(t_\varepsilon, A) &= \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon)\in A, \eta_\varepsilon(t_\varepsilon)=1\}\\ &\le \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1}\le t_\varepsilon\}\to 0 \text{ as } \varepsilon\to 0. \end{aligned} \tag{3.126}$$

Due to the symmetricity of conditions **F–J** and $\mathbf{N}_1$ with respect to the indices $i, j = 1, 2$, the ergodic relations, analogous to the mentioned above ergodic relations for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ and $P_{\varepsilon,21}(t_\varepsilon, A)$, also take place for probabilities $P_{\varepsilon,22}(t_\varepsilon, A)$ and $P_{\varepsilon,v_\varepsilon,12}(t_\varepsilon, A)$. □

3.5.2 Time Compression Factors v_ε and w_ε

Let introduce function $w_\varepsilon = (p_{\varepsilon,12} + p_{\varepsilon,21})^{-1}$. This function possess useful asymptotic properties different of asymptotic properties of function $v_\varepsilon = p_{\varepsilon,12}^{-1} + p_{\varepsilon,21}^{-1}$.

The following lemma present some useful relations between functions v_ε and w_ε.

Lemma 3.3 *If conditions* $\mathbf{M}_1$ *holds and condition* $\mathbf{K}_\beta$ *holds, for some* $\beta\in[0,\infty]$. *Then,* $0 < w_\varepsilon < v_\varepsilon < \infty$, $\varepsilon\in[01]$, *and*:

(i) *If* $\beta\in(0,\infty)$, *then* $v_\varepsilon \sim p_{\varepsilon,12}^{-1}(1+\beta) \sim p_{\varepsilon,21}^{-1}(1+\beta^{-1})$ *as* $\varepsilon\to 0$, *while* $w_\varepsilon \sim p_{\varepsilon,12}^{-1}(1+\beta^{-1})^{-1} \sim p_{\varepsilon,21}^{-1}(1+\beta)^{-1}$ *as* $\varepsilon\to 0$, *and, thus,* $w_\varepsilon \sim \frac{\beta}{(1+\beta)^2}v_\varepsilon$ *as* $\varepsilon\to 0$.

(ii) *If* $\beta = 0$, *then* $v_\varepsilon \sim p_{\varepsilon,12}^{-1}$ *as* $\varepsilon\to 0$, *while* $w_\varepsilon \sim p_{\varepsilon,21}^{-1} \prec v_\varepsilon$ *as* $\varepsilon\to 0$.

(iii) *If* $\beta = \infty$, *then* $v_\varepsilon \sim p_{\varepsilon,21}^{-1}$ *as* $\varepsilon\to 0$, *while* $w_\varepsilon \sim p_{\varepsilon,12}^{-1} \prec v_\varepsilon$ *as* $\varepsilon\to 0$.

Here and henceforth, symbols $f'_\varepsilon \sim f''_\varepsilon$ as $\varepsilon \to 0$ and $f'_\varepsilon \prec f''_\varepsilon$ as $\varepsilon \to 0$ are used for two functions $0 < f'_\varepsilon, f''_\varepsilon \to \infty$ as $\varepsilon \to 0$ in the sense that, respectively, $f'_\varepsilon / f''_\varepsilon \to 1$ as $\varepsilon \to 0$ and $f'_\varepsilon / f''_\varepsilon \to 0$ as $\varepsilon \to 0$.

Proposition **(i)** of Lemma 3.3 implies that, in the case, where condition condition $\mathbf{K}_\beta$ holds, for some $\beta \in (0, \infty)$, relations $t_\varepsilon / v_\varepsilon \to t$ and $t_\varepsilon / w_\varepsilon \to t$ as $\varepsilon \to 0$ generate, for every $t \in [0, \infty]$, equivalent, in some sense, asymptotic time zones.

Propositions **(ii)** and **(iii)** of Lemma 3.3 imply that, in the case, where condition $\mathbf{K}_0$ or $\mathbf{K}_\infty$ holds, relations $t_\varepsilon / v_\varepsilon \to t$ and $t_\varepsilon / w_\varepsilon \to t$ as $\varepsilon \to 0$ generate, for every $t \in [0, \infty]$, essentially different asymptotic time zones.

We should assume in this case that "short" times $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ satisfy, additionally to the asymptotic relation (3.114), the following asymptotic relation,

$$t_\varepsilon / w_\varepsilon \to t \in [0, \infty] \text{ as } \varepsilon \to 0. \tag{3.127}$$

3.5.3 *Short Time Ergodic Theorems for Singularly Perturbed Alternating Regenerative Processes - II*

In this subsection, we consider the case, where parameter $t = \infty$ in relation (3.127). In this case, relations (3.114) and (3.127) mean that,

$$w_\varepsilon \prec t_\varepsilon \prec v_\varepsilon \text{ as } \varepsilon \to 0. \tag{3.128}$$

The corresponding limiting probabilities take the following forms, for $A \in \mathcal{B}_{\mathbb{X}}$, $i, j = 1, 2$,

$$\pi^{(0)}_{0,j}(A) = \begin{cases} \pi_{0,1}(A) & \text{for } j = 1, \\ 0 & \text{for } j = 2. \end{cases} \tag{3.129}$$

and

$$\pi^{(\infty)}_{0,j}(A) = \begin{cases} 0 & \text{for } j = 1, \\ \pi_{0,2}(A) & \text{for } j = 2. \end{cases} \tag{3.130}$$

It is useful to note that the above limiting probabilities $\pi^{(0)}_{0,j}(A)$ and $\pi^{(\infty)}_{0,j}(A)$ coincide with the corresponding limiting probabilities for semi-regularly perturbed alternating regenerative processes, respectively, with parameter $\beta = 0$, given in relation (3.41), and $\beta = \infty$, given in relation (3.42).

The following theorems take place.

Theorem 3.11 *Let conditions* **F**–**J**, $\mathbf{N}_1$, *and* $\mathbf{K}_0$ *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *and* $t_\varepsilon / v_\varepsilon \to 0$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi^{(0)}_{0,j}(A) \text{ as } \varepsilon \to 0. \tag{3.131}$$

Theorem 3.12 *Let conditions* **F–J**, $\mathbf{N}_1$, *and* $\mathbf{K}_\infty$ *holds. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *and* $t_\varepsilon / v_\varepsilon \to 0$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,j}^{(\infty)}(A) \text{ as } \varepsilon \to 0. \tag{3.132}$$

Proof First, let us prove Theorem 3.11.

It can be noted that the analysis of asymptotic behaviour for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ can be performed in absolutely analogous way with those presented in relations (3.117)–(3.123), in the proof of Theorem 3.10. The only difference is that parameter $\beta = 0$, and, thus, the limiting random variable in the analogue of asymptotic relation (3.119) has the form, $e_{0,1}\zeta$, where ζ is a random variable exponentially distributed, with parameter 1. This analysis yields that the following asymptotic relation takes place, for every $A \in \Gamma$ and any $t_\varepsilon / v_\varepsilon \to 0$ as $\varepsilon \to 0$,

$$P_{\varepsilon,11}(t_\varepsilon, A) \to \pi_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.133}$$

The asymptotic behaviour for probabilities $P_{\varepsilon,21}(t_\varepsilon, A)$ differs in this case of those presented in Theorem 3.10. As a matter of fact, the asymptotic relation analogous to (3.124) does not take place.

In this case, random variables $v_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \xrightarrow{d} 0$ as $\varepsilon \to 0$. This asymptotic relation does not imply relation analogous to (3.120). The right normalising function for random variables $\tilde{\tau}_{\varepsilon,1}$ is, in this case, $w_\varepsilon \sim p_{\varepsilon,21}^{-1}$ as $\varepsilon \to 0$. According this relation and relation (3.64), if $\eta_\varepsilon(0) = 2$, then,

$$w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \xrightarrow{d} e_{0,2}\zeta \text{ as } \varepsilon \to 0, \tag{3.134}$$

where ζ is a random variable exponentially distributed, with parameter 1.

Probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ and $P_{\varepsilon,21}(t_\varepsilon, A)$ are connected by the following renewal type relation,

$$\begin{aligned} P_{\varepsilon,21}(t_\varepsilon, A) &= \int_0^{t_\varepsilon} P_{\varepsilon,11}(t_\varepsilon - s, A)\mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \in ds\} \\ &= \int_0^{t_\varepsilon / w_\varepsilon} P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, A)\mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \in ds\} \\ &= \int_0^{\infty} P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, A)\mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \in ds\}, \end{aligned} \tag{3.135}$$

where function $P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, A)$ is defined as 0 for $t_\varepsilon - sw_\varepsilon < 0$.

Let us take an arbitrary $s_\varepsilon \to s \in [0, \infty)$ as $\varepsilon \to 0$. Obviously, $(t_\varepsilon - s_\varepsilon w_\varepsilon)/w_\varepsilon = t_\varepsilon / w_\varepsilon - s_\varepsilon \to \infty$ and, thus, $(t_\varepsilon - s_\varepsilon w_\varepsilon) \to \infty$ as $\varepsilon \to 0$.

Also, $(t_\varepsilon - s_\varepsilon w_\varepsilon)/v_\varepsilon = t_\varepsilon / v_\varepsilon - s_\varepsilon w_\varepsilon / v_\varepsilon \to 0$ as $\varepsilon \to 0$.

That is why, according relation (3.133), the following asymptotic relation take place, for $A \in \Gamma$ and $s \in [0, \infty)$,

$$P_{\varepsilon,11}(t_\varepsilon - s_\varepsilon w_\varepsilon, A) \to \pi_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.136}$$

Relations (3.134) and (3.136) imply, by Lemma 3.2 given in Sect. 3.4.3 that the following relation takes place, for $A \in \Gamma$,

$$P_{\varepsilon,21}(t_\varepsilon, A) \to \int_0^\infty \pi_{0,1}(A)\mathsf{P}\{e_{0,2}\zeta \in ds\} = \pi_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.137}$$

As was pointed out in Sect. 3.2.5, the phase space $\mathbb{X} \in \Gamma$. Also, $\pi_{0,1}(\mathbb{X}) = 1$. Thus, relations (3.133) and (3.137) imply that the following relation holds, for $A \in \Gamma$ and $i = 1, 2$,

$$\begin{aligned} P_{\varepsilon,i2}(t_\varepsilon, A) &\le P_{\varepsilon,i2}(t_\varepsilon, \mathbb{X}) = 1 - P_{\varepsilon,i1}(t_\varepsilon, \mathbb{X}) \\ &\to 1 - \pi_{0,1}(\mathbb{X}) = 0 \text{ as } \varepsilon \to 0. \end{aligned} \tag{3.138}$$

The proof of Theorem 3.11 is completed.

The proof of Theorem 3.12 is absolutely analogous to the proof of Theorem 3.11, due to simmetrisity conditions **F**–**J** and $\mathbf{N}_1$ with respect to indices $i, j = 1, 2$. The only formula (3.129) for the corresponding limiting probabilities should be replaced by formula (3.130). □

3.5.4 *Short Time Ergodic Theorems for Singularly Perturbed Alternating Regenerative Processes - III*

In this subsection, we consider the case, where parameter $t \in (0, \infty)$, in relation (3.127). In this case, relation (3.127) means that,

$$t_\varepsilon \sim t w_\varepsilon \text{ as } \varepsilon \to 0, \text{ where } t \in (0, \infty). \tag{3.139}$$

According propositions **(ii)** and **(iii)** of Lemma 3.3, if condition $\mathbf{K}_0$ or $\mathbf{K}_\infty$, then $w_\varepsilon \prec v_\varepsilon$ as $\varepsilon \to 0$, and, thus, relation (3.139) implies that a "short" time relation (3.114) holds, i.e., $t_\varepsilon / v_\varepsilon \to 0$ as $\varepsilon \to 0$.

The corresponding limiting probabilities take the following forms, for $A \in \Gamma$, $i, j = 1, 2$ and $t \in (0, \infty)$,

$$\dot{\pi}^{(0)}_{0,ij}(t, A) = \begin{cases} \pi_{0,1}(A) & \text{for } i = 1, j = 1, \\ 0 & \text{for } i = 1, j = 2, \\ (1 - e^{-t/e_{0,2}})\pi_{0,1}(A) & \text{for } i = 2, j = 1, \\ e^{-t/e_{0,2}}\pi_{0,2}(A) & \text{for } i = 2, j = 2. \end{cases} \tag{3.140}$$

and

$$\dot{\pi}^{(\infty)}_{0,ij}(t, A) = \begin{cases} e^{-t/e_{0,1}}\pi_{0,1}(A) & \text{for } i = 1, j = 1, \\ (1 - e^{-t/e_{0,1}})\pi_{0,2}(A) & \text{for } i = 1, j = 2, \\ 0 & \text{for } i = 2, j = 1, \\ \pi_{0,2}(A) & \text{for } i = 2, j = 2. \end{cases} \tag{3.141}$$

The following theorems take place.

Theorem 3.13 *Let conditions* **F–J**, $\mathbf{N}_1$ *and* $\mathbf{K}_0$ *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and any* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \dot{\pi}^{(0)}_{ij}(t, A) \text{ as } \varepsilon \to 0. \tag{3.142}$$

Theorem 3.14 *Let conditions* **F–J**, $\mathbf{N}_1$ *and* $\mathbf{K}_\infty$ *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \dot{\pi}^{(\infty)}_{0,ij}(t, A) \text{ as } \varepsilon \to 0. \tag{3.143}$$

Proof First, let us prove Theorem 3.13.

It can be noted, as in the proof of Theorem 3.11, that the analysis of asymptotic behaviour for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ can be performed in absolutely analogous way with those presented in relations (3.117)–(3.123), in the proof of Theorem 3.10. The only difference is that parameter $\beta = 0$, and, thus, the limiting random variable in the analogue of asymptotic relation (3.119) has the form, $e_{0,1}\zeta$, where ζ is a random variable exponentially distributed, with parameter 1.

This analysis yields that the asymptotic relation (3.133) takes place, i.e., $P_{\varepsilon,11}(t_\varepsilon, A) \to \pi_{0,1}(A)$ as $\varepsilon \to 0$, for $A \in \Gamma$ and any $t_\varepsilon / v_\varepsilon \to 0$ as $\varepsilon \to 0$.

Also, as in the proof of Theorem 3.11, the renewal type relation (3.135), connecting probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ and $P_{\varepsilon,21}(t_\varepsilon, A)$, takes place.

Let us take an arbitrary $s_\varepsilon \to s \in [0, \infty)$ as $\varepsilon \to 0$. Obviously, $(t_\varepsilon - s_\varepsilon w_\varepsilon)/w_\varepsilon = t_\varepsilon / w_\varepsilon - s_\varepsilon \to t - s$ as $\varepsilon \to 0$.

Thus, for $t > s$, the following relations holds, $(t_\varepsilon - s_\varepsilon w_\varepsilon) = (t_\varepsilon / w_\varepsilon - s_\varepsilon) w_\varepsilon \to \infty$ as $\varepsilon \to 0$ and $(t_\varepsilon - s_\varepsilon w_\varepsilon)/v_\varepsilon = (t_\varepsilon / w_\varepsilon - s_\varepsilon) w_\varepsilon / v_\varepsilon \to 0$ as $\varepsilon \to 0$.

Also, for $t < s$ function $(t_\varepsilon - s_\varepsilon w_\varepsilon) = (t_\varepsilon / w_\varepsilon - s_\varepsilon) w_\varepsilon \to -\infty$ for $\varepsilon \to 0$.

That is why, according relation (3.133) and the definition of $P_{\varepsilon,11}(t_\varepsilon - s w_\varepsilon, A) = 0$, for $t_\varepsilon - s w_\varepsilon < 0$, in relation (3.135), the following asymptotic relation holds, for $A \in \Gamma$ and $s \ne t$,

$$P_{\varepsilon,11}(t_\varepsilon - s_\varepsilon w_\varepsilon, A) \to \pi_{0,1}(A)\mathrm{I}(t > s) \text{ as } \varepsilon \to 0. \tag{3.144}$$

Note that convergence of $P_{\varepsilon,11}(t_\varepsilon - s_\varepsilon w_\varepsilon, A)$ as $\varepsilon \to 0$ is not guarantied for $s = t$. However, the distribution of limiting ransom variable in relation (3.134) is exponential and, thus, it has not an atom at any point $t > 0$.

Therefore, relations (3.134) and (3.144) imply, by Lemma 3.2 given in Sect. 3.4.3, that the following relation takes place, for $A \in \Gamma$ and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such

that $t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ as $\varepsilon \to 0$,

$$P_{\varepsilon,21}(t_\varepsilon, A) = \int_0^\infty P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, A)\mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \in ds\},$$
$$\to \int_0^\infty \pi_{0,1}(A)\mathrm{I}(t > s)e_{0,2}^{-1}e^{-s/e_{0,2}}ds$$
$$= (1 - e^{-t/e_{0,2}})\pi_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.145}$$

It remains to give the asymptotic analysis of asymptotic behaviour for probabilities $P_{\varepsilon,12}(t_\varepsilon, A)$ and $P_{\varepsilon,22}(t_\varepsilon, A)$.

As was pointed out in Sect. 3.2.5, the phase space $\mathbb{X} \in \Gamma$. Also, $\pi_{0,1}(\mathbb{X}) = 1$. Thus, relation (3.133) implies that the following relation holds, for $A \in \Gamma$ and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that $t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ as $\varepsilon \to 0$,

$$P_{\varepsilon,12}(t_\varepsilon, A) \le P_{\varepsilon,12}(t_\varepsilon, \mathbb{X}) = 1 - P_{\varepsilon,11}(t_\varepsilon, \mathbb{X})$$
$$\to 1 - \pi_{0,1}(\mathbb{X}) = 0 \text{ as } \varepsilon \to 0. \tag{3.146}$$

Let us introduce random variables $\mu_{\varepsilon,2,n} = \kappa_{\varepsilon,2,n}\mathrm{I}(\eta_{\varepsilon,2,n} = 2)$, $n = 1, 2, \ldots$. Let now consider the random sequence of triplets $\langle \bar{\xi}_{\varepsilon,2,n} = \langle \xi_{\varepsilon,2,n}(t), t \ge 0\rangle, \kappa_{\varepsilon,2,n}, \mu_{\varepsilon,2,n}\rangle$, $n = 1, 2, \ldots$, the regenerative process $\xi_{\varepsilon,2}(t) = \xi_{\varepsilon,2,n}(t - \tau_{\varepsilon,2,n-1})$, for $t \in [\tau_{\varepsilon,2,n-1}, \tau_{\varepsilon,2,n})$, $n = 1, 2, \ldots$, with regeneration times $\tau_{\varepsilon,2,n} = \kappa_{\varepsilon,2,1} + \cdots + \kappa_{\varepsilon,2,n}$, $n = 1, 2, \ldots$, $\tau_{\varepsilon,2,0} = 0$, and the random lifetime $\mu_{\varepsilon,2,+} = \tau_{\varepsilon,2,\nu_{\varepsilon,2}}$, where $\nu_{\varepsilon,2} = \min(n \ge 1 : \mu_{\varepsilon,2,n} < \kappa_{\varepsilon,2,n}) = \min(n \ge 1 : \eta_{\varepsilon,2,n} = 1)$.

Let us also denote $P_{\varepsilon,2,+}(t, A) = \mathsf{P}_2\{\xi_{\varepsilon,2}(t) \in A, \mu_{\varepsilon,2,+} > t\}$. In this case, the distribution functions $\bar{F}_{\varepsilon,2}(t) = \mathsf{P}\{\kappa_{\varepsilon,2,1} \le t\}$ and $F_{\varepsilon,2}(t) = \mathsf{P}\{\kappa_{\varepsilon,2,1} \le t, \mu_{\varepsilon,2,1} \ge \kappa_{\varepsilon,2,1}\} = \mathsf{P}\{\kappa_{\varepsilon,2,1} \le t, \eta_{\varepsilon,2,1} = 2\}$, the stopping probability

$$f_{\varepsilon,2} = \mathsf{P}\{\mu_{\varepsilon,2,1} < \kappa_{\varepsilon,2,1}\} = \mathsf{P}\{\eta_{\varepsilon,2,1} = 1\} = p_{\varepsilon,21},$$

and the expectations $\bar{e}_{\varepsilon,2} = \mathsf{E}\kappa_{\varepsilon,2,1} = e_{\varepsilon,21} + e_{\varepsilon,22}$ and

$$e_{\varepsilon,2} = \mathsf{E}\kappa_{\varepsilon,2,1}\mathrm{I}(\mu_{\varepsilon,2,1} \ge \kappa_{\varepsilon,2,1}) = \mathsf{E}\kappa_{\varepsilon,2,1}\mathrm{I}(\eta_{\varepsilon,2,1} = 2) = e_{\varepsilon,22}.$$

The following relation obviously takes place, for $A \in \mathcal{B}_{\mathbb{X}}$, $t \ge 0$,

$$P_{\varepsilon,2,+}(t, A) = \mathsf{P}\{\xi_{\varepsilon,2}(t) \in A, \mu_{\varepsilon,2,+} > t\}$$
$$= \mathsf{P}_2\{\xi_\varepsilon(t) \in A, \tilde{\tau}_{\varepsilon,1} > t\}. \tag{3.147}$$

Conditions **F**–**J**, $\mathbf{N}_1$ and $\mathbf{K}_0$ imply that conditions **A**–**D** holds. Thus, conditions of Theorem 3.13 imply that all conditions of Theorem 3.3 hold for the regenerative processes $\xi_{\varepsilon,2}(t)$, $t \ge 0$ with regenerative times $\tau_{\varepsilon,2,n}$, $n = 1, 2, \ldots$ and the regenerative lifetime $\mu_{\varepsilon,2,+}$.

Therefore, the following relation holds, for any $A \in \Gamma$, and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that $p_{\varepsilon,21} t_\varepsilon = t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ as $\varepsilon \to 0$,

$$\begin{aligned} \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} &= P_{\varepsilon,2,+}(t_\varepsilon, A) \\ &\to e^{-t/e_{0,2}} \pi_{0,2}(A) \text{ as } \varepsilon \to 0. \end{aligned} \tag{3.148}$$

Probabilities $P_{\varepsilon,22}(t_\varepsilon, A)$ and $P_{\varepsilon,12}(t_\varepsilon, A)$ are connected by the following renewal type equation,

$$\begin{aligned} P_{\varepsilon,22}(t_\varepsilon, A) &= \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} \\ &\quad + \int_0^{t_\varepsilon} P_{\varepsilon,12}(t_\varepsilon - s, A) \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \in ds\} \end{aligned} \tag{3.149}$$

The integral at the right hand side of the above relation can be represented in the following form,

$$\begin{aligned} &\int_0^{t_\varepsilon} P_{\varepsilon,12}(t_\varepsilon - s, A) \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \in ds\} \\ &\quad = \int_0^{\infty} P_{\varepsilon,12}(t_\varepsilon - s w_\varepsilon, A) \mathsf{P}_2\{w_\varepsilon^{-1} \tilde{\tau}_{\varepsilon,1} \in ds\}, \end{aligned} \tag{3.150}$$

where function $P_{\varepsilon,12}(t_\varepsilon - s w_\varepsilon, A)$ is defined as 0 for $t_\varepsilon - s w_\varepsilon < 0$.

Analogously to relation (3.144), one can get using relation (3.146) and the definition of $P_{\varepsilon,12}(t_\varepsilon - s w_\varepsilon, A) = 0$, for $t_\varepsilon - s w_\varepsilon < 0$, in relation (3.150), the following asymptotic relation holds, for any $s_\varepsilon \to s \in [0, \infty)$ as $\varepsilon \to 0$, $A \in \Gamma$ and $s \ne t$,

$$P_{\varepsilon,12}(t_\varepsilon - s_\varepsilon w_\varepsilon, A) \to 0 \text{ as } \varepsilon \to 0. \tag{3.151}$$

Therefore, relations (3.134) and (3.151) imply, by Lemma 3.2 given in Sect. 3.4.3, that the following relation takes place, for $A \in \Gamma$ and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that $t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ as $\varepsilon \to 0$,

$$\begin{aligned} &\int_0^{\infty} P_{\varepsilon,12}(t_\varepsilon - s w_\varepsilon, A) \mathsf{P}_2\{w_\varepsilon^{-1} \tilde{\tau}_{\varepsilon,1} \in ds\} \\ &\quad \to \int_0^{\infty} 0 \cdot e_{0,2}^{-1} e^{-s/e_{0,2}} ds = 0 \text{ as } \varepsilon \to 0. \end{aligned} \tag{3.152}$$

Relations (3.148)–(3.150) and (3.152) imply that the following relation holds for $A \in \Gamma$ and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that $t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ as $\varepsilon \to 0$,

$$P_{\varepsilon,22}(t_\varepsilon, A) \to e^{-t/e_{0,2}} \pi_{0,2}(A) \text{ as } \varepsilon \to 0. \tag{3.153}$$

The proof of Theorem 3.13 is completed.

The proof of Theorem 3.14 is absolutely analogous to the proof of Theorem 3.13, due to simmetrisity conditions **F**–**J** and $\mathbf{N}_1$ with respect to indices $i, j = 1, 2$. The only formula (3.140) for the corresponding limiting probabilities should be replaced by formula (3.141). □

3.5.5 *Short Time Ergodic Theorems for Singularly Perturbed Alternating Regenerative Processes - IV*

In this subsection, we consider the case, where parameter $t = 0$ in relation (3.127). In this case, relation (3.127) means, for times $t_\varepsilon \to \infty$ as $\varepsilon \to 0$, that,

$$t_\varepsilon \prec w_\varepsilon \text{ as } \varepsilon \to 0. \tag{3.154}$$

The corresponding limiting probabilities are the same for both cases, where condition $\mathbf{K}_0$ or $\mathbf{K}_\infty$ holds, and take the following form, for $A \in \Gamma, i, j = 1, 2$,

$$\pi_{0,ij}(A) = \begin{cases} \pi_{0,i}(A) & \text{for } j = i, \\ 0 & \text{for } j \neq i. \end{cases} \tag{3.155}$$

The following theorem takes place.

Theorem 3.15 *Let conditions* **F**–**J**, $\mathbf{N}_1$ *and* $\mathbf{K}_0$ *or* $\mathbf{K}_\infty$ *hold. Then, for every* $A \in \Gamma, i, j = 1, 2$, *and* $0 \leq t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to 0$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,ij}(A) \text{ as } \varepsilon \to 0. \tag{3.156}$$

Proof Let us, first, assume that condition $\mathbf{K}_0$ holds.

Relation $t_\varepsilon / w_\varepsilon \to 0$ as $\varepsilon \to 0$ implies relation $t_\varepsilon / v_\varepsilon \to 0$ as $\varepsilon \to 0$. This makes it possible to repeat the part of proof of Theorem 3.10 given in relations (3.117)–(3.123) and to get the asymptotic relation,

$$\mathsf{P}_{11}(t_\varepsilon, A) \to \pi_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.157}$$

In the case, where condition $\mathbf{K}_0$ holds, $w_\varepsilon \sim p_{\varepsilon,21}^{-1}$ as $\varepsilon \to 0$. According this relation and relation (3.64), if $\eta_\varepsilon(0) = 2$, then,

$$w_\varepsilon^{-1} \tilde{\tau}_{\varepsilon,1} \xrightarrow{d} e_{0,2}\zeta \text{ as } \varepsilon \to 0, \tag{3.158}$$

where ζ is a random variable exponentially distributed, with parameter 1.

Since, we assumed that $t_\varepsilon / w_\varepsilon \to 0$ as $\varepsilon \to 0$, the above convergence in distribution relation, obviously, implies that,

$$\mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} = \mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} > t_\varepsilon w_\varepsilon^{-1}\} \to 1 \text{ as } \varepsilon \to 0. \tag{3.159}$$

If $\eta_\varepsilon(0) = 2$, then, for every $t > 0$, event $\{\eta_\varepsilon(t) = 1\} \subseteq \{\tilde{\tau}_{\varepsilon,1} \le t\}$. Thus, for every $A \in \Gamma$,

$$\begin{aligned}\mathsf{P}_{21}(t_\varepsilon, A) &= \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 1\} \\ &\le \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \le t_\varepsilon\} \to 0 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.160}$$

As was pointed out in Sect. 3.2.5, the phase space $\mathbb{X} \in \Gamma$. Also, $\pi_{0,1}(\mathbb{X}) = 1$. Thus, relation (3.157) implies that the following relation holds, for $A \in \Gamma$ and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that $t_\varepsilon / w_\varepsilon \to 0$ as $\varepsilon \to 0$,

$$\begin{aligned}P_{\varepsilon,12}(t_\varepsilon, A) &\le P_{\varepsilon,12}(t_\varepsilon, \mathbb{X}) = 1 - P_{\varepsilon,11}(t_\varepsilon, \mathbb{X}) \\ &\to 1 - \pi_{0,1}(\mathbb{X}) = 0 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.161}$$

Finally, let us analyse the asymptotic behaviour of probabilities $P_{\varepsilon,22}(t_\varepsilon, A)$ and, thus, assume that $\eta_\varepsilon(0) = 2$.

We return back to the initial alternating regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \ge 0$ with regeneration times $\tau_{\varepsilon,n}$, $n = 0, 1, \ldots$.

Recall the stopping time $\tilde{\tau}_{\varepsilon,1}$, which is the time of first hitting sate 1 by process $\eta_\varepsilon(t)$.

Let us again consider the regenerative process $\xi_{\varepsilon,2}(t)$, $t \ge 0$ with regeneration times $\tau_{\varepsilon,2,n}$, $n = 0, 1, \ldots$, and the random lifetime $\mu_{\varepsilon,2,+}$ introduced in Sect. 3.4.4.

It is readily seen that, for every $t \ge 0$,

$$\tilde{Q}_{\varepsilon,21}(t) = \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \le t\} = \mathsf{P}\{\mu_{\varepsilon,2,+} \le t\} \tag{3.162}$$

and, for every $A \in \mathcal{B}_{\mathbb{X}}$, $t \ge 0$,

$$\mathsf{P}_2\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 2, \tilde{\tau}_{\varepsilon,1} > t\} = \mathsf{P}\{\xi_{\varepsilon,2}(t) \in A, \mu_{\varepsilon,2,+} > t\}. \tag{3.163}$$

Since, $t_\varepsilon / w_\varepsilon \to 0$ as $\varepsilon \to 0$, relations (3.158) and (3.162) imply that,

$$\begin{aligned}\mathsf{P}\{\mu_{\varepsilon,2,+} > t_\varepsilon\} &= \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} \\ &= \mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} > t_\varepsilon w_\varepsilon^{-1}\} \to 1 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.164}$$

Relations (3.118) and (3.164) imply that

$$\begin{aligned}&\mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 2\} \\ &\quad - \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 2, \tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} \\ &\le \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \le t_\varepsilon\} \to 0 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.165}$$

and, analogously,

$$\begin{aligned}&\mathsf{P}\{\xi_{\varepsilon,2}(t_\varepsilon)\in A\}\\&\quad-\mathsf{P}\{\xi_{\varepsilon,2}(t_\varepsilon)\in A,\,\mu_{\varepsilon,2,+}>t_\varepsilon\}\\&\quad\le\mathsf{P}\{\mu_{\varepsilon,2,+}\le t_\varepsilon\}\to 0\text{ as }\varepsilon\to 0.\end{aligned}\tag{3.166}$$

These relations and Theorem 3.1, which can be applied to the regenerative processes $\xi_{\varepsilon,2}(t)$, imply that, for every $A\in\Gamma$,

$$\begin{aligned}\lim_{\varepsilon\to 0}\mathsf{P}_{22}(t_\varepsilon,A)&=\lim_{\varepsilon\to 0}\mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon)\in A,\,\eta_\varepsilon(t_\varepsilon)=2\}\\&=\lim_{\varepsilon\to 0}\mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon)\in A,\,\eta_\varepsilon(t_\varepsilon)=2,\,\tilde\tau_{\varepsilon,1}>t_\varepsilon\}\\&=\lim_{\varepsilon\to 0}\mathsf{P}\{\xi_{\varepsilon,2}(t_\varepsilon)\in A,\,\mu_{\varepsilon,2,+}>t_\varepsilon\}\\&=\lim_{\varepsilon\to 0}\mathsf{P}\{\xi_{\varepsilon,2}(t_\varepsilon)\in A\}=\pi_{0,2}(A).\end{aligned}\tag{3.167}$$

In the case of holding condition $\mathbf{K}_\infty$, the proof is analogous. □

3.6 Ergodic Theorems for Super-Singularly Perturbed Alternating Regenerative Processes

In this section, we present ergodic theorems for super-singularly perturbed alternating regenerative processes. As for singularly perturbed alternating regenerative processes, these theorems take different forms of super-long, long and short time ergodic theorems for different asymptotic time zones.

3.6.1 Super-Singularly Perturbed Alternating Regenerative Processes

Let us consider the alternating regenerative processes with the super-singular perturbation model, where, additionally to conditions **F–J**, the following condition holds:

$\mathbf{N}_2$: **(a)** $p_{\varepsilon,12}=0$, for $\varepsilon\in[0,1]$, and $0<p_{\varepsilon,21}\to p_{0,21}=0$ as $\varepsilon\to 0$, or **(b)** $0<p_{\varepsilon,12}\to p_{0,12}=0$ as $\varepsilon\to 0$, and $p_{\varepsilon,21}=0$, for $\varepsilon\in[0,1]$.

In this case, $v_\varepsilon=\infty$, $\varepsilon\in(0,1]$.

The role of time scaling factor is played function w_ε, $\varepsilon\in(0,1]$.

Note that $w_\varepsilon=p_{\varepsilon,21}^{-1}$, $\varepsilon\in(0,1]$, if condition $\mathbf{N}_2$ **(a)** holds, while $w_\varepsilon=p_{\varepsilon,12}^{-1}$, $\varepsilon\in(0,1]$, if condition $\mathbf{N}_2$ **(b)** holds.

We shall investigate asymptotic behaviour of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ under for $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ such that the following time scaling relation holds,

$$t_\varepsilon / w_\varepsilon \to t \in [0, \infty] \text{ as } \varepsilon \to 0. \tag{3.168}$$

It is readily seen that that conditions $\mathbf{N}_2$ **(a)** and $\mathbf{N}_2$ **(b)** are, in some sense, stronger forms, respectively, of conditions $\mathbf{K}_0$ and $\mathbf{K}_\infty$. That is why, it is expectable that the corresponding individual ergodic theorems for super-singularly perturbed alternating regenerative processes should take forms analogous to those presented for singularly perturbed alternating regenerative processes in short time ergodic Theorems 3.11–3.15, for models with asymptotic time zones generated by the asymptotic relation (3.168).

We also include in the class of super-singularly perturbed alternating regenerative processes the extremal case of absolutely singular perturbed alternating regenerative processes. This is the case, where, additionally to **F**–**J**, the following condition holds:

$\mathbf{N}_3$: $p_{\varepsilon,12}, p_{\varepsilon,21} = 0$, for $\varepsilon \in [0, 1]$.

3.6.2 *Super-Long Time Ergodic Theorems for Super-Singularly Perturbed Alternating Regenerative Processes*

In this subsection, we investigate asymptotic behaviour for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ for times $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ satisfying the following relation,

$$t_\varepsilon / w_\varepsilon \to \infty \text{ as } \varepsilon \to 0. \tag{3.169}$$

The corresponding limiting probabilities take the following form, for $A \in \Gamma$, $i, j = 1, 2$,

$$\pi^{(0)}_{0,j}(A) = \begin{cases} \pi_{0,1}(A) & \text{for } j = 1, \\ 0 & \text{for } j = 2. \end{cases} \tag{3.170}$$

and

$$\pi^{(\infty)}_{0,j}(A) = \begin{cases} 0 & \text{for } j = 1, \\ \pi_{0,2}(A) & \text{for } j = 2. \end{cases} \tag{3.171}$$

The following theorems takes place.

Theorem 3.16 *Let conditions* **F**–**J** *and* $\mathbf{N}_2$ **(a)** *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to \infty$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi^{(0)}_{0,j}(A) \text{ as } \varepsilon \to 0. \tag{3.172}$$

Theorem 3.17 *Let conditions* **F–J** *and* $\mathbf{N}_2$ **(b)** *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to \infty$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi^{(\infty)}_{0,j}(A) \text{ as } \varepsilon \to 0. \tag{3.173}$$

Proof The asymptotic behaviour for probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ is obviously given by Theorems 3.1. Indeed, if $\eta_\varepsilon(0) = 1$, then condition $\mathbf{N}_2$ **(a)** implies that the process $\xi_\varepsilon(t)$, $t \ge 0$ coincides with the process $\xi_{\varepsilon,1}(t)$, $t \ge 0$, while the process $\eta_\varepsilon(t) = 1$, $t \ge 0$. Thus, the following relation takes place, for any $A \in \Gamma$, and any $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$,

$$P_{\varepsilon,11}(t_\varepsilon, A) \to \pi_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.174}$$

Also, for any $A \in \Gamma$ and $t \ge 0$,

$$P_{\varepsilon,12}(t, A) = 0. \tag{3.175}$$

According relation (3.64), if $\eta_\varepsilon(0) = 2$, then,

$$w_\varepsilon^{-1} \tilde{\tau}_{\varepsilon,1} \xrightarrow{d} e_{0,2}\zeta \text{ as } \varepsilon \to 0, \tag{3.176}$$

where ζ is a random variable exponentially distributed, with parameter 1.

Probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ and $P_{\varepsilon,21}(t_\varepsilon, A)$ are connected by the following renewal type relation,

$$\begin{aligned} P_{\varepsilon,21}(t_\varepsilon, A) &= \int_0^{t_\varepsilon} P_{\varepsilon,11}(t_\varepsilon - s, A)\mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \in ds\} \\ &= \int_0^{\infty} P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, A)\mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \in ds\}, \end{aligned} \tag{3.177}$$

where function $P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, A)$ is defined as 0 for $t_\varepsilon - sw_\varepsilon < 0$.

Let us take an arbitrary $s_\varepsilon \to s \in [0, \infty)$ as $\varepsilon \to 0$. Obviously, $(t_\varepsilon - s_\varepsilon w_\varepsilon)/w_\varepsilon = t_\varepsilon / w_\varepsilon - s_\varepsilon \to \infty$ as $\varepsilon \to 0$. That is why, according relation (3.174), the following asymptotic relation take place, for $A \in \Gamma$ and $s \in [0, \infty)$,

$$P_{\varepsilon,11}(t_\varepsilon - s_\varepsilon w_\varepsilon, A) \to \pi_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.178}$$

Relations (3.176) and (3.178) imply, by Lemma 3.2 given in Sect. 3.4.3 that the following relation takes place, for $A \in \Gamma$,

$$P_{\varepsilon,21}(t_\varepsilon, A) \to \int_0^{\infty} \pi_{0,1}(A)\mathsf{P}\{e_{0,2}\zeta \in ds\} = \pi_{0,1}(A) \text{ as } \varepsilon \to 0. \tag{3.179}$$

As was pointed out in Sect. 3.2.5, the phase space $\mathbb{X} \in \Gamma$. Also, $\pi_{0,1}(\mathbb{X}) = 1$. Thus, relations (3.179) implies that the following relation holds, for $A \in \Gamma$,

$$P_{\varepsilon,22}(t_\varepsilon, A) \le P_{\varepsilon,22}(t_\varepsilon, \mathbb{X}) = 1 - P_{\varepsilon,21}(t_\varepsilon, \mathbb{X}) \to 1 - \pi_{0,1}(\mathbb{X}) = 0 \text{ as } \varepsilon \to 0. \tag{3.180}$$

The proof of Theorem 3.16 is completed.
The proof of Theorem 3.17 is absolutely analogous. □

3.6.3 *Long Time Ergodic Theorems for Super-Singularly Perturbed Alternating Regenerative Processes*

In this subsection, we investigate asymptotic behaviour for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ for times $0 \le t_\varepsilon \to \infty$ as $\varepsilon \to 0$ satisfying the following relation,

$$t_\varepsilon / w_\varepsilon \to t \in (0, \infty) \text{ as } \varepsilon \to 0. \tag{3.181}$$

The corresponding limiting probabilities take that following form for $A \in \Gamma$, $i, j = 1, 2$ and $t \in (0, \infty)$,

$$\dot{\pi}^{(0)}_{0,ij}(t, A) = \begin{cases} \pi_{0,1}(A) & \text{for } i = 1, j = 1, \\ 0 & \text{for } i = 1, j = 2, \\ (1 - e^{-t/e_{0,2}})\pi_{0,1}(A) & \text{for } i = 2, j = 1, \\ e^{-t/e_{0,2}}\pi_{0,2}(A) & \text{for } i = 2, j = 2. \end{cases} \tag{3.182}$$

and

$$\dot{\pi}^{(\infty)}_{0,ij}(t, A) = \begin{cases} e^{-t/e_{0,1}}\pi_{0,1}(A) & \text{for } i = 1, j = 1, \\ (1 - e^{-t/e_{0,1}})\pi_{0,2}(A) & \text{for } i = 1, j = 2, \\ 0 & \text{for } i = 2, j = 1, \\ \pi_{0,2}(A) & \text{for } i = 2, j = 2. \end{cases} \tag{3.183}$$

The following theorems take place.

Theorem 3.18 *Let conditions* **F–J** *and* $\mathbf{N}_2$ **(a)** *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \dot{\pi}^{(0)}_{ij}(t, A) \text{ as } \varepsilon \to 0. \tag{3.184}$$

Theorem 3.19 *Let conditions* **F–J** *and* $\mathbf{N}_2$ **(b)** *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \le t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to t \in (0, \infty)$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \dot{\pi}^{(\infty)}_{0,ij}(t, A) \text{ as } \varepsilon \to 0. \tag{3.185}$$

Proof The asymptotic behaviour of probabilities $P_{\varepsilon,1j}(t_\varepsilon, A)$, $j = 1, 2$ is given, under the assumption that condition $\mathbf{N}_2$ **(a)** holds, is given by relations (3.174) and (3.175), in the proof of Theorem 3.16.

Recall again relation (3.64). If $\eta_\varepsilon(0) = 2$, then, for $u \geq 0$,

$$\mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \leq u\} \to 1 - e^{-u/e_{0,2}} \text{ as } \varepsilon \to 0. \tag{3.186}$$

Also recall the renewal type relation connecting probabilities $P_{\varepsilon,11}(t_\varepsilon, A)$ and $P_{\varepsilon,21}(t_\varepsilon, A)$,

$$\begin{aligned} P_{\varepsilon,21}(t_\varepsilon, A) &= \int_0^{t_\varepsilon} P_{\varepsilon,11}(t_\varepsilon - s, A)\mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \in ds\} \\ &= \int_0^{\infty} P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, , A)\mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \in ds\}, \end{aligned} \tag{3.187}$$

where function $P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, A)$ is defined as 0 for $t_\varepsilon - sw_\varepsilon < 0$.

Let us take an arbitrary $s_\varepsilon \to s \in [0, \infty)$ as $\varepsilon \to 0$. Obviously, $(t_\varepsilon - s_\varepsilon w_\varepsilon)/w_\varepsilon = t_\varepsilon/w_\varepsilon - s_\varepsilon \to t - s$ as $\varepsilon \to 0$. That is why, according relation (3.174) and the above definition of $P_{\varepsilon,11}(t_\varepsilon - sw_\varepsilon, A) = 0$, for $t_\varepsilon - sw_\varepsilon < 0$. the following asymptotic relation holds, for $A \in \Gamma$ and $s \neq t$,

$$P_{\varepsilon,11}(t_\varepsilon - s_\varepsilon w_\varepsilon, A) \to \pi_{0,1}(A)\mathrm{I}(t > s) \text{ as } \varepsilon \to 0. \tag{3.188}$$

Note that convergence of $P_{\varepsilon,11}(t_\varepsilon - s_\varepsilon w_\varepsilon, A)$ as $\varepsilon \to 0$ is not guarantied for $s = t$. However the limiting distribution in relation (3.186) is exponential and, thus, it has not an atom at any point $t > 0$.

Therefore, relations (3.186) and (3.188) imply, by Lemma 3.2 given Sect. 3.4.3, that the following relation takes place, for $A \in \Gamma$ and $t \in (0, \infty)$,

$$\begin{aligned} P_{\varepsilon,21}(t_\varepsilon, A) &\to \int_0^{\infty} \pi_{0,1}(A)\mathrm{I}(t > s)e_{0,2}^{-1}e^{-s/e_{0,2}}ds \\ &= (1 - e^{-t/e_{0,2}})\pi_{0,1}(A) \text{ as } \varepsilon \to 0. \end{aligned} \tag{3.189}$$

It remains to give the asymptotic analysis of asymptotic behaviour for probabilities $P_{\varepsilon,22}(t_\varepsilon, A)$.

Let us introduce random variables $\mu_{\varepsilon,2,n} = \kappa_{\varepsilon,2,n}\mathrm{I}(\eta_{\varepsilon,2,n} = 2)$, $n = 1, 2, \ldots$. Let now consider the random sequence of triplets $\langle\bar{\xi}_{\varepsilon,2,n} = \langle\xi_{\varepsilon,2,n}(t), t \geq 0\rangle, \kappa_{\varepsilon,2,n}, \mu_{\varepsilon,2,n}\rangle$, $n = 1, 2, \ldots$, the regenerative process $\xi_{\varepsilon,2}(t) = \xi_{\varepsilon,2,n}(t - \tau_{\varepsilon,2,n-1})$, for $t \in [\tau_{\varepsilon,2,n-1}, \tau_{\varepsilon,2,n})$, $n = 1, 2, \ldots$, with regeneration times $\tau_{\varepsilon,2,n} = \kappa_{\varepsilon,2,1} + \cdots + \kappa_{\varepsilon,2,n}$, $n = 1, 2, \ldots, \tau_{\varepsilon,2,0} = 0$, and the random lifetime $\mu_{\varepsilon,2,+} = \tau_{\varepsilon,2,\nu_{\varepsilon,2}}$, where $\nu_{\varepsilon,2} = \min(n \geq 1 : \mu_{\varepsilon,2,n} < \kappa_{\varepsilon,2,n}) = \min(n \geq 1 : \eta_{\varepsilon,2,n} = 1)$.

Let us also denote $P_{\varepsilon,2,+}(t, A) = \mathsf{P}_2\{\xi_{\varepsilon,2}(t) \in A, \mu_{\varepsilon,2,+} > t\}$. In this case, the distribution functions $\bar{F}_{\varepsilon,2}(t) = \mathsf{P}\{\kappa_{\varepsilon,2,1} \leq t\}$ and $F_{\varepsilon,2}(t) = \mathsf{P}\{\kappa_{\varepsilon,2,1} \leq t, \mu_{\varepsilon,2,1} \geq \kappa_{\varepsilon,2,1}\} = \mathsf{P}\{\kappa_{\varepsilon,2,1} \leq t, \eta_{\varepsilon,2,1} = 2\}$, the stopping probability

$$f_{\varepsilon,2} = \mathsf{P}\{\mu_{\varepsilon,2,1} < \kappa_{\varepsilon,2,1}\} = \mathsf{P}\{\eta_{\varepsilon,2,1} = 1\} = p_{\varepsilon,21},$$

and the expectations $\bar{e}_{\varepsilon,2} = \mathsf{E}\kappa_{\varepsilon,2,1} = e_{\varepsilon,21} + e_{\varepsilon,22}$ and

$$e_{\varepsilon,2} = \mathsf{E}\kappa_{\varepsilon,2,1}\mathrm{I}(\mu_{\varepsilon,2,1} \geq \kappa_{\varepsilon,2,1}) = \mathsf{E}\kappa_{\varepsilon,2,1}\mathrm{I}(\eta_{\varepsilon,2,1} = 2) = e_{\varepsilon,22}.$$

Condition $\mathbf{N}_2$ **(a)** implies that, for every $A \in \mathcal{B}_{\mathbb{X}}$, $t \geq 0$,

$$\begin{aligned} P_{\varepsilon,22}(t, A) &= \mathsf{P}_2\{\xi_\varepsilon(t) \in A, \tilde{\tau}_{\varepsilon,1} > t\} \\ &= \mathsf{P}\{\xi_{\varepsilon,2}(t) \in A, \mu_{\varepsilon,2,+} > t\} = P_{\varepsilon,2,+}(t, A). \end{aligned} \tag{3.190}$$

Conditions **F**–**J** and $\mathbf{N}_2$ **(a)** and imply that conditions **A**–**D** holds. Thus, conditions of Theorem 3.18 imply that all conditions of Theorem 3.3 hold for the regenerative processes $\xi_{\varepsilon,2}(t)$, $t \geq 0$ with regenerative times $\tau_{\varepsilon,2,n}$, $n = 1, 2, \ldots$ and random lifetimes $\mu_{\varepsilon,2,+}$. Therefore, the following relation holds, for any $A \in \Gamma$, and $t_\varepsilon \to t \in (0, \infty)$ as $\varepsilon \to 0$,

$$P_{\varepsilon,22}(t_\varepsilon, A) = P_{\varepsilon,2,+}(t_\varepsilon, A) \to e^{-t/e_{0,2}}\pi_{0,2}(A) \text{ as } \varepsilon \to 0. \tag{3.191}$$

The proof of Theorem 3.18 is completed.
The proof of Theorem 3.19 is absolutely analogous. □

3.6.4 *Short Time Ergodic Theorems for Super-Singularly Perturbed Alternating Regenerative Processes*

In this subsection, we investigate asymptotic behaviour for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ for times $0 \leq t_\varepsilon \to \infty$ as $\varepsilon \to 0$ satisfying the following relation,

$$t_\varepsilon / w_\varepsilon \to 0 \text{ as } \varepsilon \to 0. \tag{3.192}$$

The corresponding limiting probabilities are the same for both case, where condition $\mathbf{N}_2$ **(a)** or $\mathbf{N}_2$ **(b)** holds. They take the following form, for $A \in \Gamma$, $i, j = 1, 2$,

$$\pi_{0,ij}(A) = \begin{cases} \pi_{0,i}(A) & \text{for } j = i, \\ 0 & \text{for } j \neq i. \end{cases} \tag{3.193}$$

The following theorem takes place.

Theorem 3.20 *Let conditions* **F**–**J** *and* $\mathbf{N}_2$ *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and* $0 \leq t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$ *such that* $t_\varepsilon / w_\varepsilon \to 0$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,ij}(A) \text{ as } \varepsilon \to 0. \tag{3.194}$$

Proof Let us, first, assume that condition $\mathbf{N}_2$ **(a)** holds.

The asymptotic behaviour of probabilities $P_{\varepsilon,1j}(t_\varepsilon, A)$, $j = 1, 2$ is given, under the assumption that condition $\mathbf{N}_2$ **(a)** holds, by relations (3.174) and (3.175), in the proof of Theorem 3.17.

It is readily seen that, for every $t \geq 0$,

$$\tilde{Q}_{\varepsilon,21}(t) = \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \leq t\} = \mathsf{P}\{\mu_{\varepsilon,2,+} \leq t\} \tag{3.195}$$

and, for every $A \in \mathcal{B}_{\mathbb{X}}$, $t \geq 0$,

$$\mathsf{P}_2\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = 2, \tilde{\tau}_{\varepsilon,1} > t\} = \mathsf{P}\{\xi_{\varepsilon,2}(t) \in A, \mu_{\varepsilon,2,+} > t\}. \tag{3.196}$$

According relation (3.64), if $\eta_\varepsilon(0) = 2$, random variables, $w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} \xrightarrow{d} e_{0,2}\zeta$ as $\varepsilon \to 0$, where ζ is a random variable exponentially distributed with parameter 1. Since, we assumed that $t_\varepsilon/w_\varepsilon \to 0$ as $\varepsilon \to 0$, the above convergence in distribution relation and relation (3.195) imply that,

$$\begin{aligned}\mathsf{P}\{\mu_{\varepsilon,2,+} > t_\varepsilon\} &= \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} \\ &= \mathsf{P}_2\{w_\varepsilon^{-1}\tilde{\tau}_{\varepsilon,1} > t_\varepsilon w_\varepsilon^{-1}\} \to 1 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.197}$$

Relations (3.196) and (3.197) imply that

$$\begin{aligned}&\mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 2\} - \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 2, \tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} \\ &\quad \leq \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \leq t_\varepsilon\} \to 0 \text{ as } \varepsilon \to 0,\end{aligned} \tag{3.198}$$

and, analogously,

$$\begin{aligned}&\mathsf{P}\{\xi_{\varepsilon,2}(t_\varepsilon) \in A\} - \mathsf{P}\{\xi_{\varepsilon,2}(t_\varepsilon) \in A, \mu_{\varepsilon,2,+} > t_\varepsilon\} \\ &\quad \leq \mathsf{P}\{\mu_{\varepsilon,2,+} \leq t_\varepsilon\} \to 0 \text{ as } \varepsilon \to 0,\end{aligned} \tag{3.199}$$

These relations and Theorem 3.1, which can be applied to the regenerative processes $\xi_{\varepsilon,2}(t)$, imply that, for every $A \in \Gamma$,

$$\begin{aligned}\lim_{\varepsilon\to 0} \mathsf{P}_{22}(t_\varepsilon, A) &= \lim_{\varepsilon\to 0} \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 2\} \\ &= \lim_{\varepsilon\to 0} \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 2, \tilde{\tau}_{\varepsilon,1} > t_\varepsilon\} \\ &= \lim_{\varepsilon\to 0} \mathsf{P}\{\xi_{\varepsilon,2}(t_\varepsilon) \in A, \mu_{\varepsilon,2,+} > t_\varepsilon\} \\ &= \lim_{\varepsilon\to 0} \mathsf{P}\{\xi_{\varepsilon,2}(t_\varepsilon) \in A\} = \pi_{0,2}(A).\end{aligned} \tag{3.200}$$

If $\eta_\varepsilon(0) = 2$, then, for every $t > 0$, event $\{\eta_\varepsilon(t) = 1\} \subseteq \{\tilde{\tau}_{\varepsilon,1} \leq t\}$. Thus, for every $A \in \Gamma$,

$$\begin{aligned}\mathsf{P}_{21}(t_\varepsilon, A) &= \mathsf{P}_2\{\xi_\varepsilon(t_\varepsilon) \in A, \eta_\varepsilon(t_\varepsilon) = 1\} \\ &\leq \mathsf{P}_2\{\tilde{\tau}_{\varepsilon,1} \leq t_\varepsilon\} \to 0 \text{ as } \varepsilon \to 0.\end{aligned} \tag{3.201}$$

The proof for the case, where condition $\mathbf{N}_2$ **(b)** holds, is absolutely analogous to the above proof, due to the simmetrisity conditions **F–J** and $\mathbf{N}_2$ **(a)** and **(b)** with respect to indices $i, j = 1, 2$. □

3.6.5 Ergodic Theorems for Absolutely Singular Perturbed Alternating Regenerative Processes

This is the extremal and trivial case, where condition $\mathbf{N}_3$ holds.

In this case, the process $\xi_\varepsilon(t), t \geq 0$ coincides with the process $\xi_{\varepsilon,i}(t), t \geq 0$ and the process $\eta_\varepsilon(t) = i, t \geq 0$, if $\eta_\varepsilon(0) = i$, for $i = 1, 2$.

Thus, the asymptotic behaviour for probabilities $P_{\varepsilon,ii}(t_\varepsilon, A)$ is given by Theorem 3.1.

Also, probabilities $P_{\varepsilon,12}(t, A)$, $P_{\varepsilon,21}(t, A) = 0$, for $t \geq 0$.

The above remarks can be summarised in following theorem.

Theorem 3.21 *Let conditions* **F–J** *and* $\mathbf{N}_3$ *hold. Then, for every* $A \in \Gamma$, $i, j = 1, 2$, *and any* $0 \leq t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \pi_{0,ij}(A) \text{ as } \varepsilon \to 0. \tag{3.202}$$

3.6.6 One- and Multi-Dimensional Distributions for Perturbed Alternating Regenerative Processes

Individual ergodic theorems presented in this paper give ergodic relations for one-dimensional distributions $P_{\varepsilon,ij}(t, A) = \mathsf{P}_i\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = j\}$ for alternating regenerative processes with semi-Markov modulation $(\xi_\varepsilon(t), \eta_\varepsilon(t))$.

This makes it possible to weaken the model assumption **(j)** formulated in Sect. 3.2.4. This assumption concerns multi-dimensional joint distributions of random variables $\xi_{\varepsilon,i,n}(t_k), k = 1, \ldots, r$ and $\kappa_{\varepsilon,i,n}$, $\eta_{\varepsilon,i,n}$. This assumption can be replaced by the weaker assumption that the joint distributions of random variables $\xi_{\varepsilon,i,n}(t)$ and $\kappa_{\varepsilon,i,n}$, $\eta_{\varepsilon,i,n}$ do not depend on $n \geq 1$, for every $t \geq 0$ and $i = 1, 2$.

Process $(\xi_\varepsilon(t), \eta_\varepsilon(t), t \geq 0$ still will process a weaken, say, one-dimensional regenerative property, which, in fact, means that one-dimensional distributions $P_{\varepsilon,ij}(t, A) = \mathsf{P}_i\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = j\}, t \geq 0, i = 1, 2$ satisfy the system of renewal type Eq. (3.16). Respectively, formulations of conditions, propositions and proofs of Theorems 3.4–3.21 still remain to be valid.

3.6.7 Alternating Regenerative Processes with Transition Periods

Ergodic theorems for perturbed alternating regenerative processes can be generalised to such processes with transition periods. In this case, the model assumption **(j)** for-

mulated in Sect. 3.2.4. is assumed to hold only for $n \geq 2$. The alternating regenerative process $(\xi_\varepsilon(t), \eta_\varepsilon(t))$, $t \geq 0$ has the transition period $[0, \tau_{\varepsilon,1})$, while the shifted process $(\xi_\varepsilon^{(1)}(t), \eta_\varepsilon^{(1)}(t)) = (\xi_\varepsilon(\tau_{\varepsilon,1} + t), \eta_\varepsilon(\tau_{\varepsilon,1} + t)) \geq 0$ is a usual alternating regenerative process.

All quantities appearing in conditions **G–J** the renewal type Eq. (3.16) and relations (3.15) and (3.18) should be, in this case, defined using shifted sequence of triplets $\langle \bar{\xi}_{\varepsilon,i,2} = \langle \xi_{\varepsilon,i,2}(t), t \geq 0\rangle, \kappa_{\varepsilon,i,2}, \eta_{\varepsilon,i,2}\rangle$, $i = 1, 2$. It is also natural to index the above mentioned quantities by the upper index $^{(1)}$, for example, to use notation $P_{\varepsilon,i,j}^{(1)}(t, A) = \mathsf{P}_i\{\xi_\varepsilon^{(1)}(t) \in A, \eta_\varepsilon^{(1)}(t) = j\}$, etc. Probabilities $P_{\varepsilon,ij}^{(1)}(t, A)$ satisfy the system of renewal type Eq. (3.16). Theorems 3.4–3.21 present, in this case, the corresponding ergodic relations for these probabilities.

Instead of condition **E**, condition **G** should be assumed to hold for probabilities $\tilde{p}_{\varepsilon,ij} = \mathsf{P}\{\eta_{\varepsilon,i,1} = j\}$, $i, j = 1, 2$ and condition **H** (with omitted the non-arithmetic assumption) for transition probabilities $\tilde{Q}_{\varepsilon,ij}(t) = \mathsf{P}\{\kappa_{\varepsilon,i,1} \leq t, \eta_{\varepsilon,i,1} = j\}$, $t \geq 0, i, j = 1, 2$. The corresponding ergodic relations for probabilities $P_{\varepsilon,ij}(t_\varepsilon, A) = \mathsf{P}_i\{\xi_\varepsilon(t) \in A, \eta_\varepsilon(t) = j\}$ take the form similar with the asymptotic relation (3.21). If, for example, $P_{\varepsilon,ij}^{(1)}(t_\varepsilon, A) \to \pi_{0,ij}^{(\beta)}(t, A)$ as $\varepsilon \to 0$, for $i = 1, 2$, then,

$$P_{\varepsilon,ij}(t_\varepsilon, A) \to \tilde{p}_{0,i1}\pi_{0,1j}^{(\beta)}(t, A) + \tilde{p}_{0,i2}\pi_{0,2j}^{(\beta)}(t, A) \text{ as } \varepsilon \to 0.$$

3.7 Summary of Results

In this section, a summary of results obtained in the paper and a list of some open directions for further extension of its results are given.

3.7.1 Summary of Results

As it was pointed in the introduction, the paper presents results of complete analysis and classification of ergodic theorems for perturbed alternating regenerative processes modulated by two states semi-Markov processes.

It is shown that the forms of the corresponding ergodic relations and limiting probabilities appearing in these relations are essentially determined by two parameters.

The first one is parameter $\beta \in [0, \infty]$, which asymptotically balance switching probabilities $p_{\varepsilon,12}$ and $p_{\varepsilon,21}$ between two alternative variants of regenerative processes, in the form of asymptotic relation, $p_{\varepsilon,12}/p_{\varepsilon,21} \to \beta$ as $\varepsilon \to 0$.

The second one is a time scaling parameter $t \in [0, \infty]$, which determines the asymptotic time zones for time $t_\varepsilon \to \infty$ as $\varepsilon \to 0$, in the form of one of two asymptotic relations, $t_\varepsilon/v_\varepsilon \to t$ or $t_\varepsilon/w_\varepsilon \to t$ as $\varepsilon \to 0$, with time scaling factors, respectively, $v_\varepsilon = p_{\varepsilon,12}^{-1} + p_{\varepsilon,21}^{-1}$ or $w_\varepsilon = (p_{\varepsilon,12} + p_{\varepsilon,21})^{-1}$.

The variants of ergodic relations are presented in Theorems 3.4–3.21, which we split in groups as ergodic theorems for regularly perturbed alternating regenerative processes, and short, long, and super-long time ergodic theorems for singularly and super-singularly perturbed alternating regenerative processes.

The classification of the corresponding individual ergodic theorems is summarised in the given below Tables 3.1, 3.2 and 3.3 (where numbers of theorems, their conditions, the corresponding asymptotic time zones, and the limiting probabilities are given, respectively, in columns 1, 2, 3 and 4).

It should be noted that the limiting probabilities appearing in Theorems 3.4–3.21 have the forms $\pi_{0,j}^{(\beta)}(A) = \rho_j(\beta)\pi_{0,j}(A)$, $\pi_{0,ij}^{(\beta)}(t, A) = p_{ij}^{(\beta)}(t)\pi_{0,j}(A)$ and $\dot{\pi}_{0,ij}^{(0)}(t, A) = \dot{p}_{ij}^{(0)}(t)\pi_{0,j}(A)$, $\dot{\pi}_{0,ij}^{(\infty)}(t, A) = \dot{p}_{ij}^{(\infty)}(t)\pi_{0,j}(A)$. Coefficients $\rho_j(\beta)$ and

Table 3.1 Classification of ergodic theorems: regular perturbations

T	Conditions	Asymptotic time zones	Limiting probabilities
4	**F–J**, $\mathbf{M}_1$, $\beta = 1$	$t_\varepsilon \to \infty$	$\pi_{0,j}^{(1)}(A)$
5	**F–J**, $\mathbf{M}_2$, $\beta \in (0, \infty)$	$t_\varepsilon \to \infty$	$\pi_{0,j}^{(\beta)}(A)$
6	**F–J**, $\mathbf{M}_3$, $\beta = 0$	$t_\varepsilon \to \infty$	$\pi_{0,j}^{(0)}(A)$
7	**F–J**, $\mathbf{M}_3$, $\beta = \infty$	$t_\varepsilon \to \infty$	$\pi_{0,j}^{(\infty)}(A)$

Table 3.2 Classification of ergodic theorems: singular perturbations

T	Conditions	Asymptotic time zones	Limiting probabilities
8	**F–J**, $\mathbf{N}_1$, $\mathbf{K}_\beta$, $\beta \in [0, \infty]$	$v_\varepsilon \prec t_\varepsilon$	$\pi_{0,j}^{(\beta)}(A)$
9	**F–J**, $\mathbf{N}_1$, $\mathbf{K}_\beta$, $\beta \in [0, \infty]$	$t_\varepsilon \sim tv_\varepsilon$, $t \in (0, \infty)$	$\pi_{0,ij}^{(\beta)}(t, A)$
10	**F–J**, $\mathbf{N}_1$, $\mathbf{K}_\beta$, $\beta \in (0, \infty)$	$t_\varepsilon \prec v_\varepsilon$, $t_\varepsilon \to \infty$	$\pi_{0,ij}(A)$
11	**F–J**, $\mathbf{N}_1$, $\mathbf{K}_0$	$w_\varepsilon \prec t_\varepsilon \prec v_\varepsilon$	$\pi_{0,j}^{(0)}(A)$
12	**F–J**, $\mathbf{N}_1$, $\mathbf{K}_\infty$	$w_\varepsilon \prec t_\varepsilon \prec v_\varepsilon$	$\pi_{0,j}^{(\infty)}(A)$
13	**F–J**, $\mathbf{N}_1$, $\mathbf{K}_0$	$t_\varepsilon \sim tw_\varepsilon$, $t \in (0, \infty)$	$\dot{\pi}_{0,ij}^{(0)}(t, A)$
14	**F–J**, $\mathbf{N}_1$, $\mathbf{K}_\infty$	$t_\varepsilon \sim tw_\varepsilon$, $t \in (0, \infty)$	$\dot{\pi}_{0,ij}^{(\infty)}(t, A)$
15	**F–J**, $\mathbf{N}_1$, $\mathbf{K}_0$ or $\mathbf{K}_\infty$	$t_\varepsilon \prec w_\varepsilon$, $t_\varepsilon \to \infty$	$\pi_{0,ij}(A)$

Table 3.3 Classification of ergodic theorems: super-singular perturbations

T	Conditions	Asymptotic time zones	Limiting probabilities
16	**F–J**, $\mathbf{N}_2$ **(a)**	$w_\varepsilon \prec t_\varepsilon$	$\pi_{0,j}^{(0)}(A)$
17	**F–J**, $\mathbf{N}_2$ **(b)**	$w_\varepsilon \prec t_\varepsilon$	$\pi_{0,j}^{(\infty)}(A)$
18	**F–J**, $\mathbf{N}_2$ **(a)**	$t_\varepsilon \sim tw_\varepsilon$, $t \in (0, \infty)$	$\dot{\pi}_{0,ij}^{(0)}(t, A)$
19	**F–J**, $\mathbf{N}_2$ **(b)**	$t_\varepsilon \sim tw_\varepsilon$, $t \in (0, \infty)$	$\dot{\pi}_{0,ij}^{(\infty)}(t, A)$
20	**F–J**, $\mathbf{N}_2$	$t_\varepsilon \prec w_\varepsilon$, $t_\varepsilon \to \infty$	$\pi_{0,ij}(A)$
21	**F–J**, $\mathbf{N}_3$	$t_\varepsilon \to \infty$	$\pi_{0,ij}(A)$

$p_{ij}^{(\beta)}(t)$, $\dot{p}_{ij}^{(0)}(t)$, $\dot{p}_{ij}^{(\infty)}(t)$ can be interpreted as, respectively, stationary probabilities and transition probabilities for some semi-Markov processes or Markov chains controlling switching of regimes for the limiting alternating regenerative processes, while $\pi_{0,j}(A)$ are the stationary probabilities for these limiting regenerative processes corresponding to different regimes.

It is worth noting that limiting probabilities $\pi_{0,j}^{(\beta)}(A)$ and $\pi_{0,ij}^{(\beta)}(t, A)$, $\dot{\pi}_{0,ij}^{(0)}(t, A)$, $\dot{\pi}_{0,ij}^{(\infty)}(t, A)$ possess some natural continuity properties as functions of parameters $\beta \in [0, \infty]$ and $t \in [0, \infty]$.

In particular, the limiting probabilities $\pi_{0,j}^{(\beta)}(A)$, which appear, for regularly perturbed alternating regenerative processes, in Theorems 3.4–3.7, and, for singularly and super-singularly perturbed alternating regenerative processes, in Theorems 3.8, 3.9, 3.11, 3.12, 3.16, and 3.17, are continuous functions of parameter $\beta \in [0, \infty]$.

Analogously, the limiting probabilities $\pi_{0,ij}^{(\beta)}(t, A)$, which appear, for singularly perturbed alternating regenerative processes, in Theorem 3.9, are continuous functions of parameter $(\beta, t) \in [0, \infty] \times [0, \infty]$, except points $(0, 0)$ and $(\infty, 0)$. Also, the limiting probabilities $\dot{\pi}_{0,ij}^{(0)}(t, A)$ and $\dot{\pi}_{0,ij}^{(\infty)}(t, A)$, which appear, for singularly and super-singularly perturbed alternating regenerative processes, in Theorems 3.13, 3.14, 3.18, and 3.19, are continuous functions of parameter $t \in [0, \infty]$.

Limits, $\pi_{0,ij}^{(\beta)}(0, A) = \lim_{t \to 0} \pi_{0,ij}^{(\beta)}(t, A) = \pi_{0,ij}(A)$, for $\beta \in (0, \infty)$, while $\pi_{0,ij}^{(0)}(0, A) = \lim_{t \to 0} \pi_{0,ij}^{(0)}(t, A) = \pi_{0,j}^{(0)}(A)$ and $\pi_{0,ij}^{(\infty)}(0, A) = \lim_{t \to 0} \pi_{0,ij}^{(\infty)}(t, A) = \pi_{0,j}^{(\infty)}(A)$. Also, limit, $\pi_{0,ij}^{(\beta)}(\infty, A) = \lim_{t \to \infty} \pi_{0,ij}^{(\beta)}(t, A) = \pi_{0,j}^{(\beta)}(A)$, for $\beta \in [0, \infty]$. Here, $\pi_{0,ij}(A)$ are the limiting probabilities appearing in Theorems 3.10, 3.15, 3.20, and 3.21.

The latter asymptotic relations have a natural explanation. As a matter of fact, there exists some kind of "competition" between the velocities with which the switching probabilities $p_{\varepsilon,12}$, $p_{\varepsilon,21}$ tends to zero and time t_ε tends to infinity, for singularly and super-singularly perturbed alternating regenerative processes. Probabilities $p_{\varepsilon,12}$, $p_{\varepsilon,21}$ determine the "grade of singularity" for perturbed alternating regenerative processes. These processes become more singular if parameter $\beta_\varepsilon = p_{\varepsilon,12}/p_{\varepsilon,21}$ takes values close to 0 or ∞. The time parameter t controls the "grade of ergodicity" for perturbed alternating regenerative processes. Values of β_ε closer to 0 or ∞ and smaller values of parameter t promote convergence of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ to limiting probabilities $\pi_{0,ij}(A) = \mathrm{I}(j = i)\pi_{0,i}(A)$, characteristic for absolutely singular alternating regenerative processes (for which, switching of regimes is impossible). Moderate values of β_ε asymptotically separated of 0 and ∞ and larger values of parameter t promote manifestation of ergodic phenomena and convergence of probabilities $P_{\varepsilon,ij}(t_\varepsilon, A)$ to limiting probabilities $\pi_{0,j}^{(\beta)}(A) = \rho_j(\beta)\pi_{0,j}(A)$, which are characteristic for regular alternating regenerative processes.

3.7.2 *Directions for Future Research*

Let us list some directions for further continuation of research studies, which results are presented in the paper.

It is clear that analogous individual ergodic theorems can be obtained for perturbed alternating regenerative processes with discrete time.

Individual ergodic theorems presented in this paper relate to one-dimensional distributions of alternating regenerative processes. It would be useful to get analogous ergodic theorems for multi-dimensional distributions.

A very interesting and prospective direction for future studies is individual ergodic theorems for singularly and super-singularly perturbed multi-alternating regenerative processes. These are models analogous to those studied in the present paper, but with alternative regenerative processes choosing from some parametric finite or more general sets, which serve as the phase space for the corresponding switching (modulating) semi-Markov processes.

An important is model of alternating regenerative processes with terminating regeneration times, where the regenerative processes $\xi_{\varepsilon,i,n}(t)$, $t \geq 0$ and random vectors $(\kappa_{\varepsilon,i,n}, \eta_{\varepsilon,i,n})$ are independent.

Another important model is where the processes $\xi_{\varepsilon,i,n}(t)$, $t \geq 0$ are of Markov processes, random variables $\kappa_{\varepsilon,i,n}$ are some Markov moments for these processes, and the switching random variables $\eta_{\varepsilon,i,n}$ are determined by some events for random trajectories $\xi_{\varepsilon,i,n}(t)$, $t \in [0, \kappa_{\varepsilon,i,n})$.

An unbounded area of applications constitute queuing, reliability, control and other types of stochastic systems with alternating regimes of function.

Results in the listed above directions shall be presented in future publications.

References

1. Alimov, D., Shurenkov, V.M.: Markov renewal theorems in a scheme of series. Ukr. Mat. Zh. **21**(11), 1443–1448 (1990) (English translation in Ukr. Math. J. **42**(11), 1283–1288)
2. Anisimov, V.V.: Random processes with discrete components. Vysshaya Shkola and Izdatel'stvo Kievskogo Universiteta, Kiev, 183 pp. (1988)
3. Anisimov, V.V.: Switching Processes in Queueing Models. Applied Stochastic Methods Series, 345 pp. ISTE, London; Wiley, Hoboken (2008)
4. Asmussen, S.: Applied Probability and Queues, 2nd edn. Applications of Mathematics, Stochastic Modelling and Applied Probability, vol. 51, xii+438 pp. Springer, New York (2003)
5. Avrachenkov, K.E., Filar, J.A., Howlett, P.G.: Analytic Perturbation Theory and Its Applications, xii+372 pp. SIAM, Philadelphia (2013)
6. Bini, D.A., Latouche, G., Meini, B.: Numerical Methods for Structured Markov Chains. Numerical Mathematics and Scientific Computation, xii+327 pp. Oxford Science Publications, Oxford University Press, New York (2005)
7. Cohen, J.W.: On Regenerative Processes in Queueing Theory. Lecture Notes in Economics and Mathematical Systems, vol. 121, ix+93 pp. Springer, Berlin (1976)
8. Courtois, P.J.: Decomposability. Queueing and Computer System Applications. ACM Monograph Series, xiii+201 pp. Academic Press, New York (1977)
9. Cox, D.R.: Renewal Theory, ix+142 pp. Methuen, London; Wiley, New York (1962)
10. Englund, E.: Nonlinearly perturbed renewal equations with applications. Doctoral dissertation, Umeå University (2001)
11. Englund, E., Silvestrov, D.S.: Mixed large deviation and ergodic theorems for regenerative processes with discrete time. In: Jagers, P., Kulldorff, G., Portenko, N., Silvestrov, D. (eds.)

Proceedings of the Second Scandinavian–Ukrainian Conference in Mathematical Statistics, vol. I, Umeå (1997). Theory Stoch. Process. **3(19)**(1–2), 164–176 (1997)
12. Feller, W.: An Introduction to Probability Theory and Its Applications, Vol. II, 2nd edn. xxiv+669 pp. Wiley, New York (1971)
13. Gyllenberg, M., Silvestrov, D.S.: Nonlinearly perturbed regenerative processes and pseudo-stationary phenomena for stochastic systems. Stoch. Process. Appl. **86**, 1–27 (2000)
14. Gyllenberg, M., Silvestrov, D.S.: Quasi-Stationary Phenomena in Nonlinearly Perturbed Stochastic Systems. De Gruyter Expositions in Mathematics, vol. 44, ix+579 pp. Walter de Gruyter, Berlin (2008)
15. Iglehart, D.L., Shedler, G.S.: Regenerative Simulation of Response Times in Networks of Queues. Lecture Notes in Control and Information Sciences, vol. 26, xii+204 pp. Springer, Berlin (1980)
16. Kartashov, N.V.: Strong Stable Markov Chains, 138 pp. VSP, Utrecht and TBiMC, Kiev (1996)
17. Kingman, J.F.C.: Regenerative Phenomena. Wiley Series in Probability and Mathematical Statistics, xii+190 pp. Wiley, London (1972)
18. Koroliuk, V.S., Limnios, N.: Stochastic Systems in Merging Phase Space, xv+331 pp. World Scientific, Singapore (2005)
19. Korolyuk, V., Swishchuk, A.: Semi-Markov Random Evolutions. Mathematics and Its Applications, vol. 308, x+310 pp. Kluwer, Dordrecht (1995) (English revised edition of: Semi-Markov Random Evolutions, 254 pp. Naukova Dumka, Kiev (1992))
20. Korolyuk, V.S., Turbin, A.F.: Semi-Markov Processes and Its Applications, 184 pp. Naukova Dumka, Kiev (1976)
21. Korolyuk, V.S., Turbin, A.F.: Mathematical Foundations of the State Lumping of Large Systems. Mathematics and Its Applications, vol. 264, x+278 pp. Kluwer, Dordrecht (1993) (English edition of Mathematical Foundations of the State Lumping of Large Systems, 218 pp. Naukova Dumka, Kiev (1978))
22. Korolyuk, V.S., Korolyuk, V.V.: Stochastic Models of Systems. Mathematics and Its Applications, vol. 469, xii+185 pp. Kluwer, Dordrecht (1999)
23. Kovalenko, I.N., Kuznetsov, N.Yu., Shurenkov, V.M.: Models of Random Processes. A Handbook for Mathematicians and Engineers, 446 pp. CRC Press, Boca Raton (1996) (A revised edition of Stochastic Processes. A Handbook, 366 pp. Naukova Dumka, Kiev (1983))
24. Lindvall, T.: Lectures on the Coupling Method. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, xiv+257 pp. Wiley, New York (2002) (A revised reprint of the 1992 original)
25. Ni, Y.: Nonlinearly perturbed renewal equations: asymptotic results and applications. Doctoral dissertation, vol. 106, Mälardalen University, Västerås (2011)
26. Ni, Y., Silvestrov, D., Malyarenko, A.: Exponential asymptotics for nonlinearly perturbed renewal equation with non-polynomial perturbations. J. Numer. Appl. Math. **1**(96), 173–197 (2008)
27. Petersson, M.: Perturbed discrete time stochastic models. Doctoral dissertation, Stockholm University (2016)
28. Rolski, T., Schmidli, H., Schmidt, V., Teugels, J.: Stochastic Processes for Insurance and Finance. Wiley Series in Probability and Statistics, xviii+654 pp. Wiley, New York (1999)
29. Ross, S.M.: Introduction to Probability Models, 11th edn. xvi+767 pp. Elsevier/Academic Press, Amsterdam (2014)
30. Serfozo, R.: Basics of Applied Stochastic Processes. Probability and Its Applications, xiv+443 pp. Springer, Berlin (2009)
31. Shedler, G.S.: Regeneration and Networks of Queues. Applied Probability. A Series of the Applied Probability Trust, viii+224 pp. Springer, New York (1987)
32. Shedler, G.S.: Regenerative Stochastic Simulation. Statistical Modeling and Decision Science, x+400 pp. Academic Press, Boston (1993)
33. Shurenkov, V.M.: Transition phenomena of the renewal theory in asymptotical problems of theory of random processes 1. Mat. Sbornik **112**, 115–132 (1980) (English translation in Math. USSR: Sbornik **40**(1), 107–123)

34. Shurenkov, V.M.: Transition phenomena of the renewal theory in asymptotical problems of theory of random processes 2. Mat. Sbornik **112**, 226–241 (1980) (English translation in Math. USSR: Sbornik **40**(2), 211–225)
35. Shurenkov, V.M.: Ergodic Theorems and Related Problems, viii+96 pp. VSP, Utrecht (1998) (English edition of: Ergodic Theorems and Related Problems of Theory of Random Processes, 119 pp. Naukova Dumka, Kiev (1981))
36. Shurenkov, V.M.: Ergodic Markov Processes. Probability Theory and Mathematical Statistics, 332 pp. Nauka, Moscow (1989)
37. Silvestrov, D.: Coupling for Markov renewal processes and the rate of convergence in ergodic theorems for processes with semi-Markov switchings. Acta Appl. Math. **34**, 109–124 (1994)
38. Silvestrov, D.: Improved asymptotics for ruin probabilities. In: Silvestrov, D., Martin-Löf, A. (eds.) Modern Problems in Insurance Mathematics, Chapter 5, 37–68. EAA Series, Springer, Cham (2014)
39. Silvestrov, D., Silvestrov, S.: Asymptotic expansions for stationary distributions of perturbed semi-Markov processes. In: Silvestrov, S., Rančić, M. (eds.) Engineering Mathematics II. Algebraic, Stochastic and Analysis Structures for Networks, Data Classification and Optimization, Chapter 10. Springer Proceedings in Mathematics & Statistics, vol. 179, pp. 151–222. Springer, Cham (2016)
40. Silvestrov, D., Silvestrov, S.: Asymptotic expansions for stationary distributions of nonlinearly perturbed semi-Markov processes 1, 2. Methodol. Comp. Appl. Probab. 20 pp. (2017). Part 1: https://doi.org/10.1007/s11009-017-9605-0, Part 2: https://doi.org/10.1007/s11009-017-9607-y
41. Silvestrov, D., Silvestrov, S.: Nonlinearly Perturbed Semi-Markov Processes. Springer Briefs in Probability and Mathematical Statistics, xiv+143 pp. Springer, Cham (2017)
42. Silvestrov, D., Petersson, M., Hössjer, O.: Nonlinearly perturbed birth-death-type models. In: Silvestrov, S., Rančić, M., Malyarenko, A. (eds.) Stochastic Processes and Applications, Chapter 3. Springer proceedings in mathematics & statistics, vol. 271, Springer, Cham (2018)
43. Silvestrov, D.S.: A generalization of the renewal theorem. Dokl. Akad. Nauk Ukr. SSR, Ser. A (11), 978–982 (1976)
44. Silvestrov, D.S.: The renewal theorem in a series scheme 1. Teor. Veroyatn. Mat. Stat. **18**, 144–161 (1978) (English translation in Theory Probab. Math. Stat. **18**, 155–172)
45. Silvestrov, D.S.: The renewal theorem in a series scheme 2. Teor. Veroyatn. Mat. Stat. **20**, 97–116 (1979) (English translation in Theory Probab. Math. Stat. **20**, 113–130)
46. Silvestrov, D.S.: Semi-Markov Processes with a Discrete State Space. Library for an Engineer in Reliability, 272 pp. Sovetskoe Radio, Moscow (1980)
47. Silvestrov, D.S.: Synchronized regenerative processes and explicit estimates for the rate of convergence in ergodic theorems. Dokl. Acad. Nauk Ukr. SSR, Ser. A (11), 22–25 (1980)
48. Silvestrov, D.S.: Method of a single probability space in ergodic theorems for regenerative processes 1. Math. Oper. Stat. Ser. Optim. **14**, 285–299 (1983)
49. Silvestrov, D.S.: Method of a single probability space in ergodic theorems for regenerative processes 2. Math. Oper. Stat Ser. Optim. **15**, 601–612 (1984)
50. Silvestrov, D.S.: Method of a single probability space in ergodic theorems for regenerative processes 3. Math. Oper. Stat. Ser. Optim. **15**, 613–622 (1984)
51. Silvestrov, D.S.: Exponential asymptotic for perturbed renewal equations. Teor. Ĭmovirn. Mat. Stat. **52**, 143–153 (1995) (English translation in Theory Probab. Math. Stat. **52**, 153–162)
52. Silvestrov D.S.: Limit Theorems for Randomly Stopped Stochastic Processes. Probability and Its Applications, xiv+398 pp. Springer, London (2004)
53. Silvestrov D.S.: Nonlinearly perturbed stochastic processes and systems. In: Rykov, V., Balakrishnan, N., Nikulin, M. (eds.) Mathematical and Statistical Models and Methods in Reliability, Chapter 2, 19–38. Birkhäuser, New York (2010)
54. Silvestrov, D.S., Petersson, M.: Exponential expansions for perturbed discrete time renewal equations. In: Karagrigoriou, A., Lisnianski, A., Kleyner, A., Frenkel, I. (eds.) Applied Reliability Engineering and Risk Analysis, Chapter 23, 349–362. Probabilistic Models and Statistical Inference, Wiley, New York (2014)

55. Smith, W.L.: Regenerative stochastic processes. Proc. R. Soc. Ser. A Math. Phys. Eng. Sci. **232**, 6–31 (1955)
56. Stewart, G.W.: Matrix Algorithms. Vol. I. Basic Decompositions, xx+458 pp. SIAM, Philadelphia (1998)
57. Stewart, G.W.: Matrix Algorithms. Vol. II. Eigensystems, xx+469 pp. SIAM, Philadelphia (2001)
58. Stewart, G.W., Sun, J.G.: Matrix Perturbation Theory. Computer Science and Scientific Computing, xvi+365 pp. Academic Press, Boston (1990)
59. Thorisson, H.: Coupling, Stationarity and Regeneration. Probability and Its Applications, xiv+517 pp. Springer, New York (2000)
60. Yeleiko, Ya.I., Shurenkov, V.M.: Transient phenomena in a class of matrix-valued stochastic evolutions. Theor. Ĭmovirn. Mat. Stat. **52**, 72–76 (1995) (English translation in Theory Probab. Math. Stat. **52**, 75–79)
61. Yin, G.G., Zhang, Q.: Discrete-Time Markov Chains. Two-Time-Scale Methods and Applications. Stochastic Modelling and Applied Probability, vol. 55, xix+348 pp. Springer, New York (2005)
62. Yin, G.G., Zhang, Q.: Continuous-Time Markov Chains and Applications. A Two-Time-Scale Approach, 2nd edn. Stochastic Modelling and Applied Probability, vol. 37, xxii+427 pp. Springer, New York (2013) (An extended variant of the first 1998 edition)

Chapter 4
On Baxter Type Theorems for Generalized Random Gaussian Fields

Sergey Krasnitskiy and Oleksandr Kurchenko

Abstract Some type of Baxter sums for generalized random Gaussian fields are introduced in this work. Sufficient conditions of such a sum convergence to a non-random constant are obtained. As the examples, the behavior of Baxter sums for a class of generalized fields with independent values and for a field of fractional Brownian motion is considered.

Keywords Levy–Baxter theorems · Generalized random field · Gaussian field

4.1 Introduction

Let $\xi = \xi(t), t \in [0, 1]^d \subset R^d$ be a random field, $U_n = \{U_{n,1}, \ldots, U_{n,K(n)}\}$ for every n be a family of d–dimensional rectangles that forms a partition of $[0, 1]^d$. Let $\Delta_{n,k}\xi$ be some weighted finite difference of ξ's values at the vertices of $U_{n,k}$, $k = 1, \ldots, K(n)$, $K(n) \in N$; $K(n) \to \infty$. The sum

$$S(\xi, U_n) = \sum_{k=1}^{K(n)} (\Delta_{n,k}\xi)^2$$

is called a Baxter sum (or sum of Levy–Baxter).

Limit theorems, the content of which consists in obtaining the conditions of convergence of the Baxter sums is a rather widespread research topic of probabilistic

S. Krasnitskiy (✉)
Kyiv National University of Technology and Design, Kyiv, Ukraine
e-mail: krasnits.sm@ukr.net

O. Kurchenko
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
e-mail: olkurchenko@ukr.net

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_4

studies starting at least from the work of [1]. In this work the result of this type was proved for the standard Brownian motion. The next step in this direction was made in the work [2], where Levy's result was generalized to a broader class of Gaussian random processes. After the appearance of this work, the convergence of Baxter sums for Gaussian random processes was investigated by many authors, see, for example [3, 4]. Baxter sums for Gaussian random fields was the subject of research in pioneer works on this topic [5–7]. The results of this type, which can be attributed to the domain of stochastic analysis, at the same time, have serious applications in the statistics of random processes and fields. For example the Baxter sum method was applied to the estimation of fractional Brownian motion Hurst parameter in works [8, 9]. This method was used for the estimation of covariation function parameters for multyparameter fractional Brownian fields in the article [10]. The part of monograph [11] is devoted to the application of Baxter sums method to statistics of random processes. The estimates received by the Baxter sum method have the property of consistency. One of advantages of this method lies in the possibility to the construct of non–asymptotic confidence intervals.

Proceeding from what has been said, it seems natural to extend the method of Baxter sums to other classes of random functions, for example, to generalized random processes and fields (random functions of several arguments; the theory of generalized random functions is rather completely exposited in the third chapter of monograph [12]). We note that since the values of such a field at a point are not defined, then the Baxter sum $S(\xi, U)$ given above does not in this case make sense. Therefore, in this case the Baxter sums must be defined differently. One version of this definition was proposed and applied to the estimation of the spectral density parameters of a homogeneous generalized random field in monograph [13] and also in articles [14, 15]. In the work [16] some general variants of the Baxter sums definitions for generalized Gaussian random processes was proposed and the conditions for convergence of such sums were received. The application of the obtained results for conditions of orthogonality (singularity) of probability measures corresponding to the given process was presented.

The proposed work contains generalizations of the results [16] to the case of Gaussian generalized random functions of several real arguments (generalized random fields).

4.2 Main Results

Let K_d be the space of infinitely differentiable on R^d functions with compact supports, $\xi = \xi(\varphi), \varphi \in K_d$ be a generalized random field with zero mathematical expectation. We introduce the following notations:

$$R_+^d = (0, +\infty)^d; t = (t_1, t_2, \ldots, t_d) \in R^d; h = (h_1, h_2, \ldots, h_d) \in R^d;$$

$\{\chi_{t,h}|t \in R^d, h \in R_+^d\} \subset K_d$ — a family of infinitely differentiable functions with supports $\mathrm{supp}\chi_{\bar{t},\bar{h}}$ in d–dimensional rectangles of the form $[t, t+h] = [t_1, t_1 + h_1] \times \cdots \times [t_d, t_d + h_d]$; $b_i(n) \subset N, b_i(n) \uparrow +\infty, n \to \infty, 1 \le i \le d$ – sequences of natural numbers;

$$\chi_{k,n} := \chi_{t,h} \quad \text{for} \quad t_i = \frac{k_i}{b_i(n)},\ h_i = \frac{1}{b_i(n)},$$

$$k_i = 0, 1, \ldots, b_i(n) - 1, n \ge 1,\ k = (k_1, k_2, \ldots, k_d);$$

$$A(n) = \{k_i \in (N \cup \{0\})^d | 0 \le k_i \le b_i(n) - 1, 1 \le i \le d\}, n \ge 1.$$

We put also

$$S_n(\xi) = \sum_{k \in A(n)} (\xi, \chi_{k,n})^2,$$

$$v_n(\xi) = \sum_{k,j \in A(n)} \left(E\left((\xi, \chi_{k,n})(\xi, \chi_{j,n})\right)\right)^2. \tag{4.1}$$

We say that the random variable $S_n(\xi)$ is a Baxter sum of a generalized field for the collection of functions $\{\chi_{k,n}|0 \le k_i \le b_i(n) - 1, 1 \le i \le d\}, n \ge 1$. It is useful to note that quantity $2v_n(\xi)$ is the variance of the Baxter sum $S_n(\xi)$ for a Gaussian generalized random field with zero mean, i. e. for $n \ge 1$:

$$VarS_n(\xi) = 2v_n(\xi). \tag{4.2}$$

Theorem 4.1 *Let $\xi(\varphi), \varphi \in K_d$ be a generalized Gaussian random field with zero mathematical expectation. Then the condition*

$$v_n(\xi) \to 0, n \to \infty \tag{4.3}$$

is necessary and sufficient to have the convergence of the form

$$S_n(\xi) - ES_n(\xi) \to 0, n \to \infty \tag{4.4}$$

in the mean square. If

$$\sum_{n=1}^{\infty} v_n(\xi) < \infty,$$

then the convergence in (4.4) *takes place almost surely.*

Proof Relation (4.4) follows immediately from the equality (4.2) and the relation (4.3). The convergence of series

$$\sum_{n=1}^{\infty} v_n(\xi) = \frac{1}{2} \sum_{n=1} VarS_n(\xi)$$

implies (see, for example, [17]) that random variables $S_n(\xi) - ES_n(\xi)$ converge a.s. to 0. The very equality (4.2) is obtained from the following considerations. For mathematical expectation of the product of random variables $\eta_1, \eta_2, \eta_3, \eta_4$ having a joint Gaussian distribution with zero mean, we have (see, for example, [18]) the following equality

$$E(\eta_1\eta_2\eta_3\eta_4) = E\eta_1\eta_2 E\eta_3\eta_4 + E\eta_1\eta_3 E\eta_2\eta_4 + E\eta_1\eta_4 E\eta_2\eta_3. \tag{4.5}$$

Substituting in (4.5)

$$\eta_1 = \eta_2 = (\xi, \chi_{k,n}), \eta_3 = \eta_4 = (\xi, \chi_{j,n})$$

we obtain

$$E\left((\xi, \chi_{k,n})^2(\xi, \chi_{j,n})^2\right) = 2\left(E\left(\xi, \chi_{k,n})(\xi, \chi_{j,n})\right)\right)^2 +$$

$$+E(\xi, \chi_{k,n})^2 E(\xi, \chi_{j,n})^2,\ k, j \in A(n).$$

On the other hand, obviously we have the equality

$$E(S_n(\xi))^2 = \sum_{k,j \in A(n)} E\left((\xi, \chi_{k,n})^2(\xi, \chi_{j,n})^2\right).$$

The last two equations prove (4.2). □

Corollary 4.1 *Let $\xi(\varphi), \varphi \in K_d$ be a generalized Gaussian random field from the Theorem 4.1 and $ES_n(\xi) \to c, n \to \infty$. Then condition (3) is necessary and sufficient for convergence*

$$S_n(\xi) \to c, n \to \infty \tag{4.6}$$

in the mean square. If $\sum_{n=1}^{\infty} v_n(\xi) < \infty$, hen the convergence in (4.6) takes place almost surely.

Definition 4.1 The generalized random field is said to be a field with independent values if the random variables $(\xi, \varphi), (\xi, \psi), (\varphi, \psi \in K_d)$ are independent under the condition $(\varphi \cdot \psi)(x) = 0\, \forall x \in R^d$.

Corollary 4.2 *Let ξ be a generalized Gaussian random field with independent values, $E\xi(\varphi) = 0, \varphi \in K_d$. Denote $v_n^{(0)}(\xi) = \sum_{k \in A(n)} \left(E(\xi, \chi_{k,n})^2\right)^2$. Then the condition*

$$v_n^{(0)}(\xi) \to 0, n \to \infty$$

is necessary and sufficient for the convergence (4.4) in the square mean. If

$$\sum_{n=1}^{\infty} v_n^{(0)}(\xi) < \infty,$$

then the convergence in (4.4) *takes place with probability 1 (almost sure).*

Proof Since for $k \neq j \quad \chi_{k,n}(x)\chi_{j,n}(x) = 0, x \in R^d$, then $E\left(\xi, \chi_{k,n}\right) \times$ $\times \left(\xi, \chi_{j,n}\right) = 0$ for $k \neq j$. Therefore $v_n(\xi) = v_n^{(0)}(\xi)$. □

Example 4.1 Let $\xi_c = \xi_c(\varphi), \varphi \in K_d$ be a generalized Gaussian random field with zero mean and covariance function

$$E\xi_c(\varphi)\xi_c(\psi) = c\int_{R^d} \varphi(x)\psi(x)dx,\ \varphi,\ \psi \in K_d, c = \text{const} > 0.$$

Hence, ξ_c is a generalized random field with independent values. We introduce the functions $\chi_{t,h} \in K_d$ of the following form $\chi_{t,h} : R^d \to [0, 1]$, $\text{supp}\chi_{t,h} \subset [t_1, t_1 + h_1] \times \cdots [t_d, t_d + h_d]$, $\chi_{t,h}(x) = 1$ for $x_i \in [t_i + h_i^2, t_i + h_i - h_i^2], 1 \leq i \leq d$.

Then we have

$$E\left(\xi_c, \chi_{t,h}\right)^2 = c \cdot \int_{t,t+h} \chi_{t,h}^2(x)dx =$$

$$= ch_1 \cdots h_d \left(1 + \sum_{i=1}^{d} O(h_i)\right), h_1 \to 0, \ldots h_d \to 0.$$

Therefore

$$ES_n(\xi_c) = c \cdot \sum_{k \in A(n)} \frac{1}{b_1(n) \cdots b_d(n)} \left(1 + \sum_{i=1}^{d} O\left(\frac{1}{b_i(n)}\right)\right) =$$

$$= c + \sum_{i=1}^{d} O\left(\frac{1}{b_i(n)}\right), n \to \infty.$$

Thus, we see that $ES_n(\xi_c) \to c$ when $n \to \infty$. Further,

$$v_n^{(0)}(\xi_c) = c^2 \cdot \sum_{k \in A(n)} \left(\frac{1}{b_1(n) \cdots b_d(n)}\right)^2 \left(1 + \sum_{i=1}^{d} O\left(\frac{1}{b_i(n)}\right)\right)^2 =$$

$$= O\left(\frac{1}{b_1(n) \cdots b_d(n)}\right), n \to \infty.$$

and, by virtue of the Corollaries 1, 2,

$$S_n(\xi_c) \to c, n \to \infty$$

in square mean. If

$$\sum_{n=1}^{\infty} \frac{1}{b_1(n) \cdots b_d(n)} < \infty,$$

then we have the almost sure converges of $S_n(\xi_c)$.

Remark 4.1 Let $(\Omega, \sigma, P_1, P_2)$ be a statistical structure, i.e. Ω be an elementary events space, σ be a σ-algebra of events (subsets of Ω) P_1, P_2 be a probabilistic measures on (Ω, σ). Let $\sigma(\xi) \subset \sigma$ be a σ-algebra generated by generalized random field $\xi = \xi(\varphi), \varphi \in K_d$. The field ξ is proposed to be Gaussian one in regard to both measures P_1, P_2.

Definition 4.2 Further let E_1, E_2 be the symbols of mathematical expectations with respect to measures P_1, P_2 respectively. Denote by $v_{i,n}(\xi), i = 1, 2$ the result of substitution of the symbol E_i into the $v_n(\xi)$ expression instead of $E, i = 1, 2$. Further, denote the restrictions of measures P_1, P_2 to the σ–algebra σ by $P_{1,\xi}, P_{2,\xi}$ respectively.

Corollary 4.3 *Let random field ξ and measures P_1, P_2 satisfies the conditions of Remark 1, $E_1\xi(\varphi) = E_2\xi(\varphi) = 0, \varphi \in K_d$. Also, assume that the following conditions hold:*

1. $\sum_{n=1}^{\infty} v_{i,n}(\xi) < +\infty, i = 1, 2;$
2. $E_i S_n(\xi) \to c_i, i = 1, 2, n \to \infty;$
3. $c_1 \neq c_2.$

Then measures $P_{1,\xi}, P_{2,\xi}$ are orthogonal (singular).

Proof Let $X_i = \{\omega \in \Omega : S_n(\xi) \to c_i, n \to \infty\}, i = 1, 2$. Since Theorem 4.1, it follows that $P_{1,\xi}(X_1) = P_{2,\xi}(X_2) = 1$. But $X_1 \cap X_2 = \emptyset$. □

Example 4.2 Let measures $P_{1,\xi}, P_{2,\xi}$ correspond to the fields ξ_{c_1}, ξ_{c_2} of Example 4.1 in the manner indicated above. If $c_1 \neq c_2$, then $P_{1,\xi}, P_{2,\xi}$ are orthogonal.

Theorem 4.2 *Let a generalized Gaussian random field ξ with zero mean and function family $\chi_{t,h} \subset K_d$ satisfy the following conditions:*

1. *For sufficiently small positive $h_1, \ldots, h_d$ the function $E(\xi, \chi_{t,h})$ is continuous for $t \in [0, 1]^d$ and there exist continuous functions $g : (0, \infty)^d \to (0, \infty), u : [0, 1]^d \to (0, \infty)$, such that*

$$\frac{E(\xi, \chi_{t,h})^2}{g(h)} \to u(t), h \to 0$$

   uniformly over $t \in [0, 1]^d$;
2. $v_n^{(1)}(\xi) \stackrel{\text{def}}{=} \alpha^2(n) \sum_{(k,j) \in B(n)} \left(E\left(\xi, \chi_{k,n}\right)\left(\xi, \chi_{j,n}\right)\right)^2 \to 0, n \to \infty,$
   where

$$\alpha(n) = \frac{1}{b_1(n)\cdots b_d(n) g\left(\frac{1}{b_1(n)}, \ldots, \frac{1}{b_d(n)}\right)}, n \geq 1,$$

$$B(n) = \{(k, j) | k, j \in A(n), |k_i - j_i| \geq 2, 1 \leq i \leq d\}.$$

Then

$$\tilde{S}_n(\xi) \stackrel{\text{def}}{=} \alpha(n) \sum_{k \in A(n)} \left(\xi, \chi_{k,n}\right)^2 \to \int_{[0,1]^d} u(t)dt, n \to \infty \tag{4.7}$$

in the square mean. If the series

$$\sum_{n=1}^{\infty} v_n^{(1)}(\xi), \sum_{n=1}^{\infty} \frac{1}{b_i(n)}, 1 \leq i \leq d \tag{4.8}$$

are convergent, then convergence in (4.7) *is almost sure.*

Proof Condition 1 of Theorem 4.1 implies that

$$\begin{aligned} \tilde{S}_n(\xi) = & \frac{1}{b_1(n)\cdots b_d(n)} \sum_{k \in A(n)} \frac{E\left(\xi, \chi_{k,n}\right)^2}{g\left(\frac{1}{b_1(n)}, \ldots, \frac{1}{b_d(n)}\right)} \\ & \to \int_{[0,1]^d} u(t)dt, n \to \infty. \end{aligned} \tag{4.9}$$

Using equality (4.5), as in the proof of Theorem 4.1, we obtain the expression for the variance of a random variable $\tilde{S}_n(\xi)$:

$$Var\tilde{S}_n(\xi) = 2\alpha^2(n) \sum_{(k,j) \in A(n)} \left(E\left(\xi, \chi_{k,n}\right)\left(\xi, \chi_{j,n}\right)\right)^2.$$

We set $C(n) = \{(k, j) | k, j \in A(n), \exists i \in \{1, 2, \ldots, d\} : |k_i - j_i| \leq 1\}$. Then

$$Var\tilde{S}_n(\xi) = 2\alpha^2(n) \left(\sum_{(k,j) \in C(n)} + \sum_{(k,j) \in B(n)} \right) \left(E\left(\xi, \chi_{k,n}\right)\left(\xi, \chi_{j,n}\right)\right)^2.$$

Condition 2 of Theorem 4.2 implies that

$$2\alpha^2(n) \sum_{(k,j) \in B(n)} \left(E\left(\xi, \chi_{k,n}\right)\left(\xi, \chi_{j,n}\right)\right)^2 \to 0, n \to \infty.$$

In order to estimate $\sum_{(k,j) \in C(n)}$, we use the Cauchy-Bunyakovskii inequality:

$$\left(E\left(\xi, \chi_{k,n}\right)\left(\xi, \chi_{j,n}\right)\right)^2 \leq E\left(\xi, \chi_{k,n}\right)^2 E\left(\xi, \chi_{j,n}\right)^2, k, j \in A(n).$$

We have

$$2\alpha^2(n) \sum_{(k,j)\in C(n)} \left(E\left(\xi, \chi_{k,n}\right)\left(\xi, \chi_{j,n}\right)\right)^2 \leq$$

$$\leq 2\alpha^2(n) \sum_{(k,j)\in C(n)} E\left(\xi, \chi_{k,n}\right)^2 E\left(\xi, \chi_{j,n}\right)^2$$

$$= \frac{2}{b_1^2(n)\cdots b_d^2(n)} \sum_{(k,j)\in C(n)} \left(u\left(\frac{k_1}{b_1(n)}, \ldots, \frac{k_d}{b_d(n)}\right) + o(1)\right)$$

$$\times \left(u\left(\frac{j_1}{b_1(n)}, \ldots, \frac{j_d}{b_d(n)}\right) + o(1)\right) \stackrel{\text{def}}{=} \Delta_n.$$

Let $C = \sup_{t\in[0,1]^d} u(t) + 1$. Since the number of summands of the last sum does not exceed

$$\sum_{i=1}^{d} \frac{3b_i(n) - 2}{b_i^2(n)} b_1^2(n)\cdots b_d^2(n),$$

then for all sufficiently large n we have the inequality

$$\Delta_n \leq 3C^2 \sum_{i=1}^{d} \frac{1}{b_i(n)}.$$

Consequently,

$$2\alpha^2(n) \sum_{(k,j)\in C(n)} \left(E\left(\xi, \chi_{k,n}\right)\left(\xi, \chi_{j,n}\right)\right)^2 \to 0, n \to \infty.$$

That in combination with condition 2 ensures convergence to zero $Var\tilde{S}_n(\xi)$ under $n \to \infty$. Taking (4.9) into account, we obtain the assertion of the theorem on convergence $\tilde{S}_n(\xi)$ in the mean square. From the convergence of the series (4.8) it follows that the series of variances $\sum_{n=1}^{\infty} Var\tilde{S}_n(\xi)$ converges. Thus, the convergence relation (4.7) also holds with probability 1. □

Example 4.3 Let $\xi_H = \xi_H(t), t \in R^d$ be a Gaussian random field with zero mean and covariance function

$$B_H(s,t) = \frac{1}{2^d} \prod_{i=1}^{d} \left(|s_i|^{2H_i} + |t_i|^{2H_i} - |t_i - s_i|^{2H_i}\right), \; s, t \in R^d,$$

where $H = (H_1, \ldots, H_d) \in (0,1)^d$. In the article [19] such a random field is called an anisotropic fractional Wiener field. We consider the random field ξ_H as a generalized Gaussian random field on K_d. By η_H denote the generalized mixed partial derivative of the generalized random field η_H:

$$\eta_H = \frac{\partial^d}{\partial t_1 \dots \partial t_d} \xi_H.$$

The value of the field η_H attest function $\varphi \in K_d$ is given by the following equality

$$(\eta_H, \varphi) = (-1)^d \left(\xi_H, \frac{\partial^d}{\partial t_1 \dots \partial t_d} \varphi \right), \ \varphi \in K_d.$$

The generalized Gaussian random field η_H has zero expectation and the covariance function

$$\hat{B}(\varphi, \psi) = \int_{R^d} \int_{R^d} B_H(s,t) \frac{\partial^d \varphi(s)}{\partial s_1 \dots \partial s_d} \frac{\partial^d \psi(t)}{\partial t_1 \dots \partial t_d} dt ds.$$

As in the paper [16], we denote an infinitely differentiable on R function $\mu_{u,h} = \mu_{u,h}(\cdot)$ with parameters $u \in R, h > 0$. This function satisfies the following conditions for sufficiently small values of the parameter h:

1. $\text{supp}\mu_{u,h}(\cdot) \subset \left[u, u + \exp\left(-\frac{1}{h}\right)\right] \cup \left[u + h - \exp\left(-\frac{1}{h}\right), u + h\right]$;
2. $0 \le \mu_{u,h}(s) \le \exp\left(\frac{1}{h}\right), s \in \left[u, u + \exp\left(-\frac{1}{h}\right)\right]$;
3. $\mu_{u,h}(s) = \exp\left(\frac{1}{h}\right), s \in \left[u + \exp\left(-\frac{1}{h^2}\right), u + \exp\left(-\frac{1}{h}\right) - \exp\left(-\frac{1}{h^2}\right)\right]$;
4. The graph of the function is centrally symmetric with respect to the point $\left(u + \frac{h}{2}, 0\right)$.

We set for $t = (t_1, \dots, t_d), h = (h_1, \dots, h_d), h_i > 0, 1 \le i \le d$

$$\mu_{t,h}(s) = \prod_{i=1}^{d} \mu_{t_i,h_i}(s_i), s = (s_1, \dots, s_d) \in R^d.$$

The family of functions $\chi_{t,h}$ is defined as

$$\chi_{t,h}(s) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} \mu_{t,h}(s) ds, x = (x_1, \dots, x_d) \in R^d.$$

Due to the homogeneity of the generalized random Gaussian field η_H we have

$$E(\eta_H, \chi_{t,h})^2 = E(\eta_H, \chi_{0,h})^2 = E(\xi_H, \mu_{0,h})^2 =$$

$$= \frac{1}{2^d} \prod_{i=1}^{d} \int_0^{h_i} ds_i \int_0^{h_i} \left(s_i^{2H_i} + t_i^{2H_i} - |s_i - t_i|^{2H_i} \right) \mu_{0,h_i}(s_i) \mu_{0,h_i}(t_i) dt_i.$$

Using the direct integrating and passing $h_i \to 0, 1 \le i \le d$ in the integral given below, we get (see, also [16])

$$\frac{1}{2} \int_0^{h_i} ds_i \int_0^{h_i} \left(s_i^{2H_i} + t_i^{2H_i} - |s_i - t_i|^{2H_i} \right) \mu_{0,h_i}(s_i) \mu_{0,h_i}(t_i) dt_i = h_i^{2H_i} + o(h_i^{2H_i}).$$

Consequently, $E(\eta_H, \chi_{t,h})^2 = \prod_{i=1}^{d} \left(h_i^{2H_i} + o(h_i^{2H_i})\right), h_i \to 0, 1 \le i \le d$. Thus, condition 2 of Theorem 4.2 is satisfied for

$$u(t) = 1, g(h) = h_1^{2H_1} \cdots h_d^{2H_d}, \alpha(n) = (b_1(n))^{2H_1-1} \cdots (b_d(n))^{2H_d-1}.$$

We now turn to the verification of the second condition of the theorem. Due to the homogeneity of the Gaussian random field η_H we obtain

$$v_n^{(1)}(\eta_H) = \alpha^2(n) \sum_{(k,j)\in B(n)} \left(E\left(\eta_H, \chi_{k,n}\right)\left(\eta_H, \chi_{j,n}\right)\right)^2 =$$

$$= 2^d \alpha^2(n) \sum_{j_1=0}^{b_1(n)-3} \sum_{k_1=j_1+2}^{b_1(n)-1} \cdots \sum_{j_d=0}^{b_d(n)-3} \sum_{k_d=j_d+2}^{b_d(n)-1} \left(E\left(\eta_H, \chi_{k,n}\right)\left(\eta_H, \chi_{j,n}\right)\right)^2 =$$

$$= 2^d \alpha^2(n) \sum_{l_1=0}^{b_1(n)-1} (b_1(n) - l_1) \ldots$$

$$\ldots \sum_{l_d=2}^{b_d(n)-1} (b_d(n) - l_d) \left(E\left(\eta_H, \chi_{0,n}\right)\left(\eta_H, \chi_{l,n}\right)\right)^2,$$

where $l = (l_1, \ldots, l_d)$, $\chi_{l,n} = \chi_{l_1 h_1, \ldots, l_d h_d, h}$, $h = (h_1, \ldots, h_d)$, $h_i = \frac{1}{b_i(n)}$, $1 \le i \le d$.

Therefore,

$$E\left(\eta_H, \chi_{0,n}\right)\left(\eta_H, \chi_{l,n}\right) = E\left(\xi_H, \mu_{0,n}\right)\left(\xi_H, \mu_{l,n}\right) =$$
$$= \frac{1}{2^d} \prod_{i=1}^{d} \left(\int_0^{h_i} \mu_{0,n}^2(s_i) ds_i \int_{l_i h_i}^{(l_i+1)h_i} \left(t_i^{2H_i} + s_i^{2H_i} - (t_i - s_i)^{2H_i}\right) \mu_{l_i h_i, n}^2(t_i) dt_i \right).$$

It is shown in [16] (Example 2.2) that

$$\int_0^{h_i} \mu_{0,n}^2(s_i) ds_i \int_{l_i h_i}^{(l_i+1)h_i} \left(t_i^{2H_i} + s_i^{2H_i} - (t_i - s_i)^{2H_i}\right) \mu_{l_i h_i, n}^2(t_i) dt_i =$$

$$= O\left(\frac{h_i^{2H_i}}{(l_i - 1)^{2H_i}}\right) + O\left(\exp\left(\frac{1}{h_i} - \frac{1}{h_i^2}\right)\right), h_i \to 0$$

uniformly with respect to $l_i \in \{2, 3, \ldots, b_i(n) - 1\}$.

Thus, for some $C > 0$ and sufficiently large n, we get

$$v_n^{(1)}(\eta_H) \le C \prod_{i=1}^{d} \left(\frac{1}{b_i(n)} \sum_{l_i=2}^{b_i(n)-1} \frac{1}{(l_i-1)^{4H_i}} \right).$$

Since for $n \to \infty$

$$\sum_{l_i=2}^{b_i(n)-1} \frac{1}{(l_i-1)^{4H_i}} = \begin{cases} O\left((b_i(n))^{-4H_i+1}\right), & \text{for } 0 < H_i < \frac{1}{4}, \\ O\left(\ln b_i(n)\right), & \text{for } H_i = \frac{1}{4}, \\ O(1), & \text{for } \frac{1}{4} < H_i < 1. \end{cases}$$

then

$$v_n^{(1)}(\eta_H) = \prod_{i=1}^{d} \Big(O\left((b_i(n))^{-4H_i}\right) I_{(0,\frac{1}{4})}(H_i) + O\left(\frac{\ln b_i(n)}{b_i(n)}\right) I_{\{\frac{1}{4}\}}(H_i) +$$

$$+ O\left(\frac{1}{b_i(n)}\right) I_{(\frac{1}{4},1)}(H_i) \Big),$$

where $I_X(x)$ is the indicator of the set X, i.e. $I_X(x) = \begin{cases} 0, & x \in X, \\ 1, & x \notin X \end{cases}$. Consequently,

$$v_n^{(1)}(\eta_H) \to 0, n \to \infty,$$

and the condition 2 of Theorem 4.2 is satisfied. According to this theorem,

$$(b_1(n))^{2H_1-1} \cdots (b_d(n))^{2H_d-1} \sum_{k \in A(n)} \left(\eta_H, \chi_{k,n}\right)^2 \to 1, n \to \infty \tag{4.10}$$

in square mean. If for any $\varepsilon > 0$

$$\sum_{n=1}^{\infty} (b_i(n))^{-\varepsilon} < \infty, 1 \le i \le d, \tag{4.11}$$

then, the almost sure convergence takes place in (4.10).

For brevity, let us denote the field defined in Example 4.3 ξ_H–field. The following corollary gives a sufficient condition for the orthogonality of probability measures corresponding to ξ_H–fields.

Corollary 4.4 *Let the statistical structure $(\Omega, \sigma, P_1, P_2)$ be such that the random field ξ on K_d is the ξ_{H^1}–field with respect to the measure P_1 the and ξ_{H^2}–field with respect to the measure P_2, $H^i = \left(H_1^i, \ldots, H_d^i\right), i = 1, 2$. Denote by η the generalized mixed partial derivative of ξ: $\eta = \frac{\partial^d}{\partial t_1 \ldots \partial t_d} \xi$. Let $\sigma_\xi \subset \sigma, \sigma_\eta \subset \sigma$ be the σ–algebras generated by random fields ξ, η respectively. Denote also by $P_{1,\xi}, P_{2,\xi}, P_{1,\eta}, P_{2,\eta}$*

the restrictions of measures P_1, P_2 on the σ-algebras σ_ξ, σ_η respectively. Then if the inequality

$$\sum_{j=1}^{d} H_j^1 \neq \sum_{j=1}^{d} H_j^2 \tag{4.12}$$

holds, then we have $P_{1,\xi} \perp P_{2,\xi}$, $P_{1,\eta} \perp P_{2,\eta}$, where the sign of $\perp$ denotes the orthogonality of the measures.

Proof Since $\sigma_\eta \subset \sigma_\xi$, it is sufficient to prove the second of the indicated orthogonality relations. To this end, we choose in (4.10) the sequences $b_i(n)$ coinciding with each other $b_1(n) = \ldots = b_d(n) = b(n)$ and satisfying condition (4.11). Then we set

$$X_i = \left\{ \omega \in \Omega : b(n)^{2\sum_{j=1}^{d} H_j^i - d} \sum_{k \in A(n)} \left(\eta, \chi_{k,n}\right)^2 \to 1, n \to \infty \right\}, i = 1, 2.$$

According to relation (4.10) we have the equalities $P_{1,\eta}(X_1) = P_{2,\eta}(X_2) = 1$. But according to inequality (4.12) $X_1 \cap X_2 = \emptyset$. □

References

1. Levy, P.: Le mouvement Brownian plan. Am. J. Math. **62**, 487–550 (1940)
2. Baxter, G.: A strong limit theorem for Gaussian processes. Proc. Am. Math. Soc. **7**, 522–527 (1956)
3. Gladyshev, E.G.: A new limit theorem for stochastic processes with Gaussian increments. Teor. Veroyatn. Primen. **6**, 57–66 (1961)
4. Ryzhov, Yu.M.: One limit theorem for stationary Gaussian processes. Teor. Veroyatn. Mat. Stat. **1**, 178–188 (1970)
5. Berman, S.M.: A version of the Levy-Baxter for the increments of Brownian motion of several parameters. Proc. Am. Math. Soc. **18**, 1051–1055 (1967)
6. Krasnitskiy, S.M.: On some limit theorems for random fields with Gaussian m-order increments. Teor. Veroyatn. Mat. Stat. **5**, 71–80 (1971)
7. Arak, T.V.: On Levy-Baxter type theorems for random fields. Teor. Veroyatn. Primen. **17**, 153–160 (1972)
8. Kurchenko, O.O.: A strongly consistent estimate for the Hurst parameter of fractional Brownian motion. Teor. Imovir. Mat. Stat. **67**, 45–54 (2002)
9. Breton, J.-C., Nourdin, I., Peccati, G.: Exact confidence intervals for the Hurst parameter of a fractional Brownian motion. Electron. J. Stat. **3**, 416–425 (2009)
10. Kozachenko Yu.V., Kurchenko, O.O.: An estimate for the multiparameter FBM. Theory Stoch. Process **5**(21), 113–119 (1999)
11. Prakasa Rao, B.L.S.: Statistical Inference for Fractional Diffusion Processes. Wiley, Chichester (2010)
12. Gelfand, I.M., Vilenkin, N.Ya.: Applications of Harmonic Analysis. Equipped Hilbert Spaces. Fizmatgiz, Moscow (1961)
13. Rozanov, Yu.A.: Random Fields and Stochastic Partial Differential Equations. Nauka, Moscow (1995)

14. Goryainov, V.B.: On Levy-Baxter theorems for stochastic elliptic equations. Teor. Veroyatn. Primen. **33**, 176–179 (1988)
15. Arato, N.M.: On a limit theorem for generalized Gaussian random fields corresponding to stochastic partial differential equations. Teor. Veroyatn. Primen. **34**, 409–411 (1989)
16. Krasnitskiy, S.M., Kurchenko, O.O.: Baxter type theorems for generalized random Gaussian processes. Theory Stoch. Process. **21**(37), 45–52 (2016)
17. Lamperti, J.: Stochastic Processes. A Survey of the Mathematical Theory. Vyscha shkola, Kyiv (1983)
18. Ibragimov, I.A., Rozanov, Y.A.: Gaussian Random Processes. Nauka, Moscow (1970)
19. Kamount, A.: On the fractional anisotropic random field. Probab. Math. Stat. **16**, 85–98 (1996)

Chapter 5
Limit Theorems for Quadratic Variations of the Lei–Nualart Process

Salwa Bajja, Khalifa Es-Sebaiy and Lauri Viitasaari

Abstract Let X be a Lei–Nualart process with Hurst index $H \in (0, 1)$, Z_1 be an Hermite random variable. For any $n \geq 1$, set

$$V_n = \sum_{k=0}^{n-1} \left[n^{2H} (\Delta_k X)^2 - n^{2H} \mathbb{E}(\Delta_k X)^2 \right].$$

The aim of the current paper is to derive, in the case when the Hurst index verifies $H > 3/4$, an upper bound for the total variation distance between the laws $\mathcal{L}(Z_n)$ and $\mathcal{L}(Z_1)$, where Z_n stands for the correct renormalization of V_n which converges in distribution towards Z_1. We derive also the asymptotic behavior of quadratic variations of process X in the critical case $H = 3/4$, i.e. an upper bound for the total variation distance between the $\mathcal{L}(Z_n)$ and the Normal law.

Keywords Hermite random variable · Gaussian analysis
Malliavin calculus · Convergence in law · Berry–Esseen bounds

5.1 Introduction

Quadratic variation of a stochastic process X plays an important role in different applications. For example, the concept is important if one is interested in developing stochastic calculus with respect to X. Furthermore, quadratic variations can be used to

S. Bajja
National School of Applied Sciences—Marrakesh, Cadi Ayyad University, Marrakesh, Morocco
e-mail: salwa.bajja@gmail.com

K. Es-Sebaiy (✉)
Department of Mathematics, Faculty of Science, Kuwait University, Kuwait, Kuwait
e-mail: khalifa.essebaiy@ku.edu.kw

L. Viitasaari
Department of Mathematics and System Analysis, Aalto University School of Science,
P.O. Box 11100, 00076 Aalto, Finland
e-mail: lauri.viitasaari@aalto.fi

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_5

build estimators for the model parameters such as self-similarity index or parameter describing long range dependence which have important applications in all fields of science such as hydrology, chemistry, physics, and finance to simply name a few. For such applications, one is interested to study the convergence of the quadratic variation. Furthermore, a wanted feature is to obtain a central limit theorem which allows one to apply statistical tools developed for normal random variables.

For Gaussian processes the study of quadratic variation goes back to Lévy who studied standard Brownian motion and showed the almost sure convergence

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left[W_{\frac{k}{2^n}} - W_{\frac{k-1}{2^n}} \right]^2 = 1.$$

Later this result was extended to cover more general Gaussian processes in Baxter [1] and in Gladyshev [17] for uniformly divided partitions. General subdivisions were studied in Dudley [15] and Klein and Gine [21] where the optimal condition $o\left(\frac{1}{\log n}\right)$ for the mesh of the partition was obtained in order to obtain almost sure convergence. It is also known that for the standard Brownian motion the condition $o\left(\frac{1}{\log n}\right)$ is not only sufficient but also necessary. For details on this topic see De La Vega [13] for construction, and [25] for recent results. Functional central limit theorem for general class of Gaussian processes were studied in Perrin [31]. More recently, Kubilius and Melichov [22] defined a modified Gladyshev's estimator and the authors also studied the rate of convergence. Norvaiša [27] have extended Gladyshev's theorem to a more general class of Gaussian processes. Finally, we can mention a paper by Malukas [26] who extended the results of Norvaiša to irregular partitions, and derived sufficient conditions for the mesh in order to obtain almost sure convergence.

The case of fractional Brownian motion with Hurst index $H \in (0, 1)$ were studied in details by Gyons and Leons [18] where the authors showed that appropriately scaled first order quadratic variation (that is, the one based on differences $X_{t_k} - X_{t_{k-1}}$) converges to a Gaussian limit only if $H < \frac{3}{4}$. To overcome this problem, a generalisations of quadratic variations were used in [6, 10, 12, 20]. The most commonly used generalisation is second order quadratic variations based on differences $X_{t_{k+1}} - 2X_{t_k} + X_{t_{k-1}}$ which was studied in details in a series of papers by Begyn [2–4] with applications to fractional Brownian sheet and time-space deformed fractional Brownian motion. In particular, in [2] the sufficient condition for almost sure convergence was studied with non-uniform partitions. The central limit theorem and its functional version were studied in [3, 4] with respect to a standard uniform divided partitions. Furthermore, the authors in papers [9, 33] have studied more general variations assuming that the underlying Gaussian process have stationary increments. For another generalisation, the localised quadratic variations were introduced in [5] in order to estimate the Hurst function of multifractional Gaussian process. These results have been generalised in [11, 23].

In this paper we study quadratic variation of the Lei–Nualart process, defined precisely in Sect. 5.2. The terminology stems from the fact that in [24] the authors

proved a showed that fractional Brownian motion can be decomposed into a sum of bifractional Brownian motion and another process X, which we call the Lei–Nualart process to honor the authors who first introduced this process. The Lei–Nualart process have some interesting features. Firstly, it is already known that many interesting processes including bifractional Brownian motion, subfractional Brownian motion, and a large class of selfsimilar Gaussian processes can be decomposed in terms of fractional Brownian motion and the Lei–Nualart process with different choices of parameters [19, 24, 32]. Secondly, the Lei–Nualart process has almost surely infinitely differentiable paths. On the other hand, the quadratic variation behaves similarly as the quadratic variation of fBm.

5.2 Preliminaries

In this section we briefly recall some basic facts on Gaussian analysis and Malliavin calculus that are used in this paper. For more details on the topic we refer to [29]. We begin by giving precise definition of the Lei–Nualart process.

Definition 5.1 Let $H \in (0, 1)$, $H \neq \frac{1}{2}$. The Lei–Nualart process $(X_t^H)_{t\in[0,1]}$ on $[0, 1]$ is defined as the Wiener integral

$$X_t^H = \int_0^\infty (1 - e^{-\theta t})\theta^{-H-1/2} dW_\theta.$$

Clearly, the process X^H is centered and $X_0 = 0$. The covariance function of X^H is given by

$$C_{X^H}(t, s) = \frac{-\Gamma(2 - 2H)}{H(2H - 1)} K(s, t),$$

where

$$K(s, t) = \frac{1}{2}\left[s^{2H} + t^{2H} - (s + t)^{2H}\right], \quad s, t \in [0, 1]. \tag{5.1}$$

Moreover, X^H admits a representation

$$X_t^H = \int_0^t Y_s^H ds,$$

where

$$Y_t^H = \int_0^\infty e^{-\theta t}\theta^{H-1/2} dW_\theta.$$

In particular, this shows that X^H has absolutely continuous paths, and even infinitely differentiable paths on $\mathbb{R}_+$. For more details on the process X^H, we refer to [24, 32].

Define a Hilbert space $\mathcal{H}$ by closing the set of step functions of form $f(s) = \sum_{k=1}^{n} a_k \mathbf{1}_{(0,t_k]}(s)$ with respect to the inner product $\langle 1_{(0,t]}, 1_{(0,s]} \rangle_{\mathcal{H}} = C_{X^H}(t,s)$. Denote also by $\mathcal{H}_1$ the first chaos of X, i.e. the L^2-closure of span$\{X_t : t \in [0,1]\}$. The elements of $\mathcal{H}_1$ are centered and Gaussian. Then it is well-known that the mapping $\mathbf{1}_{(0,t]} \mapsto X_t$ extends to a linear isometry I between $\mathcal{H}$ and $\mathcal{H}_1$, and hence each element $Z \in \mathcal{H}_1$ can be identified as a Wiener integral $Z = I(h)$ for some $h \in \mathcal{H}$. The elements of $\mathcal{H}$ may be not functions but distributions. However, $\mathcal{H}$ contains the subset $|\mathcal{H}|$ of all measurable functions $f : [0,1] \to \mathbb{R}$ such that

$$\int_{[0,1]^2} |f(u)||f(v)|(u+v)^{2H+2} du dv < \infty.$$

Moreover, for $f, g \in |\mathcal{H}|$, we have

$$\langle f, g \rangle_{\mathcal{H}} = 2H(2H-1) \int_{[0,1]^2} f(u) g(v) (u+v)^{2H+2} du dv.$$

Let now $q \geq 2$ be fixed and denote by $\mathcal{H}^{\otimes q}$ (resp. $\mathcal{H}^{\odot q}$) the qth tensor product (resp. qth symmetric tensor product) of $\mathcal{H}$. If we define the qth Wiener chaos $\mathcal{H}_q$ of X as the L^2-closed linear subspace of L^2_X generated by the random variables $\{H_q(Y(h)), h \in \mathcal{H}, ||h||_{\mathcal{H}} = 1\}$, where H_q is the qth Hermite polynomial defined as $H_q(x) = (-1)^q e^{\frac{x^2}{2}} \frac{d^q}{dx^q}(e^{-\frac{x^2}{2}})$, then the mapping $I_q(h^{\otimes q}) = H_q(Y(h))$ provides a linear isometry between the symmetric tensor product $\mathcal{H}^{\odot q}$ (equipped with the modified norm $||.||_{\mathcal{H}^{\odot q}} = \sqrt{q!}||.||_{\mathcal{H}^{\otimes q}}$) and $\mathcal{H}_q$. In particular, for all $f, g \in \mathcal{H}^{\odot q}$ and $q \geq 1$ one has

$$\mathbb{E}[I_q(f) I_q(g)] = q! \langle f, g \rangle_{\mathcal{H}^{\otimes q}}.$$

Let us next introduce the Hermite process of order 2. Let $H > \frac{1}{2}$ and $t \in [0,1]$ be fixed. The sequence $(\xi_n(t))_{n \geq 1}$ defined as

$$\xi_n(t) = \frac{n}{2} \sum_{j=1}^{nt} 1^{\otimes 2}_{\left(\frac{j}{n}, \frac{j+1}{n}\right]}$$

is a Cauchy sequence in $\mathcal{H}^{\otimes 2}$. Indeed, since $H > \frac{1}{2}$, we have

$$\langle 1_{(a,b]}, 1_{(u,v]} \rangle_{\mathcal{H}} = \mathbb{E}((X_b - X_a)(X_v - X_u)) = 2H(2H-1) \int_a^b \int_u^v (x+y)^{2H+2} dx dy$$

so that for any $m \geq n$

$$\langle \xi_n(t), \xi_m(t)\rangle_{\mathcal{H}^{\otimes 2}} = H^2(2H-1)^2 nm \sum_{j=1}^{mt}\sum_{k=1}^{nt} \left(\int_{j/m}^{(j+1)/m}\int_{k/n}^{(k+1)/n} (x+y)^{2H-2}dxdy\right)^2.$$

Using the mean value theorem we hence have

$$\lim_{n,m\to\infty}\langle \xi_n(t), \xi_m(t)\rangle_{\mathcal{H}^{\otimes 2}} = H^2(2H-1)^2\int_0^t\int_0^t (x+y)^{4H-4}dxdy = c_H t^{4H-2},$$

where $c_H = \frac{H^2(2H-1)^2(2^{4H-2}-2)}{(4H-2)(4H-3)}$. Let us denote by δ_t the limit of the sequence of functions $\xi_n(t)$ in $\mathcal{H}^{\otimes 2}$. For any $f \in |\mathcal{H}|^{\otimes 2}$, we have

$$\begin{aligned}\langle \xi_n(t), f\rangle_{\mathcal{H}^{\otimes 2}} &= \frac{n}{2}\sum_{k=1}^{\lfloor nt\rfloor}\langle 1^{\otimes 2}_{(\frac{k}{n},\frac{k+1}{n}]}, f\rangle_{\mathcal{H}^{\otimes 2}}\\ &= \frac{4H^2(2H-1)^2}{2} n\sum_{k=1}^{\lfloor nt\rfloor}\int_0^1\int_{\frac{k}{n}}^{\frac{k+1}{n}}(x+y)^{2H-2}dydx\\ &\quad\times\int_0^1\int_{\frac{k}{n}}^{\frac{k+1}{n}}(u+v)^{2H-2}f(x,u)dvdu\\ &\underset{n\to\infty}{\longrightarrow} 2H^2(2H-1)^2\int_0^1 t\int_{[0,1]^2}(x+u)^{2H-2}(x+v)^{2H-2}f(v,u)dudvdx\\ &= \langle \delta_t, f\rangle_{\mathcal{H}^{\otimes 2}}.\end{aligned}$$

The Hermite process is defined as follows.

Definition 5.2 The Hermite process $Z = (Z_t)_{t\in[0,1]}$ is defined by $Z_t = I_2(\delta_t)$ for $t \in [0, 1]$.

We next recall some preliminary results in order to obtain convergence in law towards a normal random variable.

Theorem 5.1 ([30]) *Let $\{F_n\}_{n\geq 1}$ be a sequence of elements in the qth Wiener chaos such that $\mathbb{E}(F_n^2) = 1$ and let N denote a standard normal random variable. Then there exists a constant C_q depending only on q such that*

$$\sup_{x\in\mathbb{R}}|\mathbb{P}(F_n < x) - \mathbb{P}(N < x)| \leq C_q\sqrt{\mathbb{E}F_n^4 - 3}.$$

Using previous result one can study quadratic variations of general Gaussian process and obtain sufficient conditions to ensure the convergence in law towards a normal random variable together with a Berry–Esseen bound. This was studied in

[35] where general Gaussian vectors and general partitions were considered. For our purposes the following result explains how Berry–Esseen bounds can be obtained easily. The result is essentially taken as a combination of results derived in [35]. However, for convenience we present the main steps of the proof.

Theorem 5.2 *Let Y be a continuous Gaussian process and denote by V_n^Y its quadratic variation defined by $V_n^Y = \sum_{k=1}^n \left[(\Delta_k Y)^2 - \mathbb{E}\,(\Delta_k Y)^2\right]$, where $\Delta_k Y = Y_{\frac{k}{n}} - Y_{\frac{k-1}{n}}$. Set $H(n) = \max_{1\le j\le n} \sum_{k=1}^n \left|\mathbb{E}(\Delta_k Y \Delta_j Y)\right|$. Then*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{V_n^Y}{\sqrt{Var(V_n^Y)}} < x \right) - \mathbb{P}(N < x) \right| \le C \frac{H(n)}{\sqrt{Var(V_n^Y)}}.$$

Proof For fixed n, let Γ_n be the covariance matrix of the increments $\Delta_k Y$, and let λ_k be its eigenvalues. Then it is well-known that $\max_{1\le k\le n} |\lambda_k| \le H(n)$. Furthermore, by [35, Lemma 2.2] (see also proof of Theorem 2.7 of [35]) we have $\mathrm{e}\left(V_n^Y\right)^2 = 2\sum_{k=1}^n \lambda_k^2$ and $\mathrm{e}\left(V_n^Y\right)^4 = 12\left[\sum_{k=1}^n \lambda_k^2\right]^2 + 24\sum_{k=1}^n \lambda_k^4$. Hence

$$\mathrm{e}\left[\frac{V_n^Y}{\sqrt{Var(V_n^Y)}}\right]^4 = 3 + 6\frac{\sum_{k=1}^n \lambda_k^4}{\left[\sum_{k=1}^n \lambda_k^2\right]^2} \le 3 + 6\frac{H(n)^2}{\sum_{k=1}^n \lambda_k^2}.$$

The claim follows now from Theorem 5.1 together with the fact that V_n is a sequence in the *2nd* chaos. □

The following lemma gives easy way to compute the function $H(n)$ and is essentially taken from [35] (see [35, Theorem 3.3]).

Lemma 5.1 ([35]) *Let Y be a continuous Gaussian process such that the function $d(s,t) = E(X_t - X_s)^2$ is in $C^{1,1}$ outside diagonal. Furthermore, assume that $|\partial_{st} d(s,t)| = O\left(|t-s|^{2H-2}\right)$ for some $H \in (0,1)$, $H \ne \frac{1}{2}$. Then*

$$\max_{1\le j\le n} \sum_{k=1}^n \left|\mathbb{E}(\Delta_k Y \Delta_j Y)\right| \le \max_{1\le j\le n} d\left(\frac{j}{n}, \frac{j-1}{n}\right) + \left(\frac{1}{n}\right)^{1\wedge 2H}.$$

5.3 Quadratic Variation of the Lei–Nualart Process

We study quadratic variations of the process X^H defined as

$$V_n = C^2 \sum_{k=0}^{n-1} \left[n^{2H}(\Delta_k X)^2 - n^{2H}\mathbb{E}(\Delta_k X)^2\right], \tag{5.2}$$

where $C = \sqrt{\frac{H(2H-1)}{\Gamma(2-2H)}}$ and $\Delta_k X = X_{\frac{k+1}{n}} - X_{\frac{k}{n}}$. By [35, Lemma 2.2], the variance of V_n is given by

$$\mathbb{E}V_n^2 = 2C^4 n^{4H} \sum_{k,r=0}^{n-1} [\mathbb{E}(\Delta_k X \Delta_r X)]^2 . \tag{5.3}$$

Moreover, by applying (5.1) we have

$$2C^2 \mathbb{E}(\Delta_k X \Delta_r X) = n^{-2H} \alpha(k, r), \tag{5.4}$$

where

$$\alpha(k, r) = (k + r + 2)^{2H} - 2(k + r + 1)^{2H} + (k + r)^{2H}. \tag{5.5}$$

Note also that $\alpha(k, r)$ has a representation

$$\alpha(k, r) = 2H(2H - 1) \int_k^{k+1} \int_r^{r+1} (x + y)^{2H-2} dxdy =$$
$$2H(2H - 1) \int_0^1 \int_0^1 (x + y + k + r)^{2H-2} dxdy$$

which we will exploit several times. Our main result is the following.

Theorem 5.3 *Let X be the Lei–Nualart process and V_n its quadratic variations defined by (5.2). Then there exists a constant $C > 0$ such that:*

Case 1 *If $H = \frac{3}{4}$, then $\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\frac{V_n}{\sqrt{var(V_n)}} < x \right) - \mathbb{P}(N < x) \right| \leq C \frac{1}{\sqrt{n}}$, where N is a standard normal random variable.*

Case 2 *If $H > \frac{3}{4}$, then $\| V_n - Z_1 \|_2 \leq C n^{3-4H}$, where Z_t is the Hermite process.*

To prove the following theorem we may use the decomposition of the subfractional Brownian motion in [32], the convergence of V_n in the case $H \in (0, 3/4)$ is as follows.

Theorem 5.4 *Let X be the Lei–Nualart process and V_n its quadratic variations defined by (5.2). such that;*

If $H \in (0, \frac{3}{4})$, then $\lim_{n \to \infty} \frac{V_n}{\sqrt{n}} = 0$.

Remark 5.1 In the case of the fBm, and thanks to the seminal works of Breuer and Major [9], Dobrushin and Major [14], Giraitis and Surgailis [16] and Taqqu [34], it is well-known that we have, as $n \to \infty$:

Case 1 If $H \in \left(0, \frac{3}{4}\right)$, then $\frac{V_n}{\sigma_H \sqrt{n}} \xrightarrow{law} \mathcal{N}(0, 1)$.

Case 2 If $H = \frac{3}{4}$, then $\frac{V_n}{\sigma_H \sqrt{n \log(n)}} \xrightarrow{law} \mathcal{N}(0, 1)$.

Case 3 If $H > \frac{3}{4}$, then $\frac{V_n}{n^{2H-1}} \xrightarrow{law} \bar{Z} \sim$ "$Hermite\ random\ variable$".

Here, $\sigma_H > 0$ denotes an (explicit) constant depending only on H. Moreover, explicit bounds for the Kolmogorov distance between the law of V_n and the standard normal law are obtained by [28, Theorem 4.1], [8, Theorem 1.2]. The following facts happen: For some constant c_H depending only on H, we have

$$d_{Kol}\left(V_n, \mathcal{N}(0,1)\right) \leq c_H \times \begin{cases} \frac{1}{\sqrt{n}} & if \quad H \in \left(0, \frac{5}{8}\right) \\ \frac{(\log n)^{3/2}}{\sqrt{n}} & if \quad H = \frac{5}{8} \\ n^{4H-3} & if \quad H \in \left(\frac{5}{8}, \frac{3}{4}\right) \\ \frac{1}{\sqrt{\log(n)}} & if \quad H = \frac{3}{4} \end{cases},$$

where $d_{Kol}(Y, Z) = \sup_{-\infty<z<\infty} |\mathbb{P}(Y \leq z) - \mathbb{P}(Z \leq z)|$. For almost sure central limit theorems for V_n, see Bercu et al. [7].

Before proving our main result we provide some auxiliary lemmas whose proofs are postponed into the appendix.

Lemma 5.2 *Let $\alpha(x, y)$ be a function defined by (5.5). Then*

(i) We have $\alpha(x, x) \sim c_H x^{2H-2}$ as $x \to \infty$, where c_H is a constant depending only on H.
(ii) For every $0 \leq x < y$ we have $\alpha(y, y) \leq \alpha(x, y) \leq \alpha(x, x)$.

The following lemma explains the asymptotic variance of the quadratic variation V_n in different cases.

Lemma 5.3 *Let V_n be defined by (5.2). Then there exist constants c and C depending only on H such that:*

(i) If $H = \frac{3}{4}$, then $cn \leq Var(V_n) \leq Cn$.
(ii) If $\frac{3}{4} < H < 1$, then $cn^{4H-2} \leq Var(V_n) \leq Cn^{4H-2}$.

Proof (*Proof of Theorem 5.3*) The case $H = \frac{3}{4}$ follows directly from Theorem 5.2. Indeed, since $|\partial_{s,t} K(s,t)| = C(s+t)^{2H-2}$ for some unimportant constant C and $(s+t)^{2H-2} \leq |t-s|^{2H-2}$, we observe $|\partial_{s,t} K(s,t)| \leq C|t-s|^{2H-2}$. Hence by combining Lemmas 5.1 and 5.3 together with the fact $\mathrm{e}(\Delta X_k)^2 \leq Cn^{-2H}$ the claim follows immediately from Theorem 5.2.

Let us next consider the case $H > \frac{3}{4}$. For any fixed $n \geq 1$ we have

$$\begin{aligned} & n^{1-2H} \sum_{k=0}^{n-1} \left[n^{2H} C^2 \left(X_{\frac{k+1}{n}} - X_{\frac{k}{n}} \right)^2 - n^{2H} C^2 \mathbb{E}\left(X_{\frac{k+1}{n}} - X_{\frac{k}{n}} \right)^2 \right] \\ &= nC^2 \sum_{k=0}^{n-1} \left[\left(X_{\frac{k+1}{n}} - X_{\frac{k}{n}} \right)^2 - \mathbb{E}\left(X_{\frac{k+1}{n}} - X_{\frac{k}{n}} \right)^2 \right] \\ &= I_2(h_n), \end{aligned}$$

where $h_n = \frac{n}{2}\sum_{k=0}^{n-1} \delta_{\frac{k}{n}}^{\otimes 2} \in \mathcal{H}^{\odot 2}$, $\delta_{\frac{k}{n}} = 1_{[\frac{k}{n}, \frac{k+1}{n}]}$. On the other hand,

$$\begin{aligned}\frac{1}{n^{4H-2}} E[V_n^2] &= \frac{1}{2}\frac{1}{n^{4H-2}} \sum_{k,l=1}^{n} \alpha(k,l)^2 = \frac{n^2}{2}\sum_{k,l=1}^{n} \alpha\left(\frac{k}{n}, \frac{l}{n}\right)^2 \\ &= \frac{n^2}{2}\sum_{k,l=1}^{n}\left(2H(2H-1)\int_{\frac{k}{n}}^{\frac{k+1}{n}}\int_{\frac{l}{n}}^{\frac{l+1}{n}} (x+y)^{2(H-1)}dxdy\right)^2 \\ &= \frac{n^2}{2}\sum_{k,l=1}^{n}\left\langle \delta_{\frac{k}{n}}, \delta_{\frac{l}{n}}\right\rangle_{\mathcal{H}}^2 = 2\sum_{k,l=1}^{n}\left\langle \frac{n}{2}\delta_{\frac{k}{n}}, \frac{n}{2}\delta_{\frac{l}{n}}\right\rangle_{\mathcal{H}}^2 \\ &= 2 \parallel h_n \parallel_{\mathcal{H}^{\otimes 2}}^2 = \parallel h_n \parallel_{\mathcal{H}^{\odot 2}}^2 .\end{aligned}$$

Since

$$\begin{aligned}\alpha(l,k) &= ((l+1)+(k+1))^{2H} - (l+(k+1))^{2H} \\ &\quad - ((l+1)+k)^{2H} + (l+k)^{2H} \\ &= 2H(2H-1)\int_k^{k+1}\int_l^{l+1} (x+y)^{2(H-1)}dxdy,\end{aligned}$$

we observe

$$\begin{aligned}\parallel h_n \parallel_{\mathcal{H}^{\odot 2}}^2 &= \tfrac{1}{2n^{4H-2}}\textstyle\sum_{k,l=1}^{n}\left(2H(2H-1)\int_k^{k+1}\int_l^{l+1}(x+y)^{2(H-1)}dxdy\right)^2 \\ &= \tfrac{2H^2(2H-1)^2}{n^{4H-2}}\textstyle\sum_{k,l=1}^{n}\left(\int_k^{k+1}\int_l^{l+1}(x+y)^{2(H-1)}dxdy\right)^2 \\ &= 2H^2(2H-1)^2n^2\textstyle\sum_{k,l=0}^{n-1}\left(\int_{\frac{k}{n}}^{\frac{k+1}{n}}\int_{\frac{l}{n}}^{\frac{l+1}{n}}(x+y)^{2(H-1)}dxdy\right)^2. \qquad (5.6)\end{aligned}$$

Here, thanks to the mean value theorem, by letting $n \to \infty$ we observe

$$\parallel h \parallel_{\mathcal{H}^{\odot 2}}^2 = 2H^2(2H-1)^2\int_0^1\int_0^1 (x+y)^{4(H-1)}dxdy. \qquad (5.7)$$

Let now $\varphi \in |\mathcal{H}|$. We have

$$\begin{aligned}\langle h_n, \varphi^{\otimes 2}\rangle_{\mathcal{H}^{\odot 2}} &= \sum_{l=0}^{n-1}\langle \frac{n}{2}\delta_{l/n}, \varphi\rangle_{\mathcal{H}}^2 \\ &= 4H^2(2H-1)^2\frac{n}{2}\sum_{l=0}^{n-1}\left(\int_{l/n}^{(l+1)/n}\int_0^1 \phi(u)(u+v)^{2H-2}dudv\right)^2\end{aligned}$$

$$= 2H^2(2H-1)^2 n \sum_{l=0}^{n-1} \left(\int_{l/n}^{(l+1)/n} \int_0^1 \phi(u)(u+v)^{2H-2} du dv \right)^2.$$

Again by letting $n \to \infty$, we get

$$\langle h, \varphi^{\otimes 2} \rangle_{\mathcal{H}^{\otimes 2}} = 2H^2(2H-1)^2 \int_0^1 \left(\int_0^1 \phi(u)(u+v)^{2H-2} du \right)^2 dv.$$

Hence

$$\begin{aligned} \langle h, h_n \rangle_{\mathcal{H}^{\otimes 2}} &= 4H^2(2H-1)^2 \tfrac{n}{2} \textstyle\sum_{l=0}^{n-1} \int_0^1 \left(\int_{k/n}^{(k+1)/n} (u+v)^{2H-2} du \right)^2 dv \\ &= 2H^2(2H-1)^2 n \textstyle\sum_{l=0}^{n-1} \int_0^1 \left(\int_{k/n}^{(k+1)/n} (u+v)^{2H-2} du \right)^2 dv. \end{aligned} \tag{5.8}$$

Combining (5.6)–(5.8) we get

$$\begin{aligned} \| h_n - h \|^2_{\mathcal{H}^{\otimes 2}} &= E|I_2(h_n) - I_2(h)|^2 \\ &= E(I_2(h_n)^2) - 2E(I_2(h_n)I_2(h)) + E(I_2(h)^2) \\ &= \| h_n \|^2_{\mathcal{H}^{\otimes 2}} - 2\langle h_n, h \rangle_{\mathcal{H}^{\otimes 2}} + \| h \|^2_{\mathcal{H}^{\otimes 2}} \\ &= 2H^2(2H-1)^2 \sum_{k,l=0}^{n-1} n^2 \left(\int_{k/n}^{(k+1)/n} \int_{l/n}^{(l+1)/n} (x+y)^{2(H-1)} dx dy \right)^2 \\ &\quad - 4H^2(2H-1)^2 n \sum_{k,l=0}^{n-1} \int_{l/n}^{(l+1)/n} \left(\int_{k/n}^{(k+1)/n} (u+v)^{2H-2} du \right)^2 dv \\ &\quad + 2H^2(2H-1)^2 \sum_{k,l=0}^{n-1} \int_{k/n}^{(k+1)/n} \int_{l/n}^{(l+1)/n} (x+y)^{4(H-1)} dx dy \\ &= 2H^2(2H-1)^2 n^{2-4H} \sum_{k,l=0}^{n-1} \left(\int_0^1 \int_0^1 (x+y+k+l)^{2(H-1)} dx dy \right)^2 \\ &\quad - 4H^2(2H-1)^2 n^{2-4H} \sum_{k,l=0}^{n-1} \int_0^1 \left(\int_0^1 (u+v+k+l)^{2H-2} du \right)^2 dv \\ &\quad + 2H^2(2H-1)^2 n^{2-4H} \sum_{k,l=0}^{n-1} \int_0^1 \int_0^1 (x+y+k+l)^{4(H-1)} dx dy. \end{aligned}$$

Let us next study different integrals. For the first integral we have

$$\int_0^1 \int_0^1 (x+y+k+l)^{2(H-1)} dxdy = \frac{1}{2H(2H-1)} \alpha(k,l)$$
$$= \frac{1}{2H(2H-1)} \left[(l+k+2)^{2H} - 2(l+k+1)^{2H} + (l+k)^{2H} \right].$$

Hence by using Taylor's theorem repeatedly we obtain

$$\int_0^1 \int_0^1 (x+y+k+l)^{2(H-1)} dxdy = (k+l)^{2H-2} + O((k+l)^{2H-3}),$$

and thus

$$\left(\int_0^1 \int_0^1 (x+y+k+l)^{2(H-1)} dxdy \right)^2 = (k+l)^{4H-4} + O((k+l)^{4H-5}). \quad (5.9)$$

For the second integral

$$\int_0^1 \left(\int_0^1 (u+v+k+l)^{2H-2} du \right)^2 dv$$

we have, for all $x \geq 0$,

$$|(1+x)^{2H-1} - 1 - (2H-1)x| = (2H-1)(2-2H) \int_0^x du \int_0^u \frac{dv}{(1+v)^{3-2H}}$$
$$\leq (2H-1)(2-2H) \int_0^x du \int_0^u dv.$$

Hence we can write

$$\int_0^1 \left(\int_0^1 (u+v+k+l)^{2H-2} du \right)^2 dv$$
$$= \frac{1}{(2H-1)^2} \int_0^1 \left((1+v+k+l)^{2H-1} - (v+k+l)^{2H-1} \right)^2 dv$$
$$= \frac{1}{(2H-1)^2} \int_0^1 (v+k+l)^{4H-2} \left(\left(\frac{1}{v+k+l} + 1 \right)^{2H-1} - 1 \right)^2 dv$$
$$= \int_0^1 (v+k+l)^{4H-2} \left(\frac{1}{v+k+l} + R\left(\frac{1}{v+k+l} \right) \right)^2 dv$$
$$= \int_0^1 (v+k+l)^{4H-4} \left(1 + (v+k+l) R\left(\frac{1}{v+k+l} \right) \right)^2 dv,$$

where the remainder term R satisfies $|R(u)| \leq (1-H)u^2$. In particular, for any $v \in [0,1]$, we have

$$(v+k+l)\left|R\left(\frac{1}{v+k+l}\right)\right| \leq \frac{1-H}{k+l-1}.$$

Hence we deduce

$$\begin{aligned}
&\int_0^1 \left(\int_0^1 (u+v+k+l)^{2H-2}du\right)^2 dv \\
&= \int_0^1 (v+k+l)^{4H-4}(1+O(1/(k+l)))^2 dv \\
&= \frac{(k+l+1)^{4H-3}-(k+l)^{4H-3}}{4H-3}(1+O(1/(k+l))) \\
&= (k+l)^{4H-3}\frac{((1+(1/(k+l)))^{4H-3}-1)}{4H-3}(1+O(1/(k+l))) \\
&= (k+l)^{4H-3}(1/(k+l)+O(1/(k+l)^2)(1+O(1/(k+l))) \\
&= (k+l)^{4H-4}+O((k+l)^{4H-5}). \qquad (5.10)
\end{aligned}$$

Finally, for the last integral we have

$$\int_0^1\int_0^1 (x+y+k+l)^{4(H-1)}dxdy = \frac{1}{(4H-3)(4H-2)}h(k,l),$$

where $h(k,l) = (k+l+2)^{4H-2} - 2(k+l+1)^{4H-2} + (k+l)^{4H-2}$. Using Taylor's theorem again we deduce $h(k,l) = (4H-2)(4H-3)(k+l)^{4H-4} + O((k+l)^{4H-5})$ and thus

$$\int_0^1\int_0^1 (x+y+k+l)^{4(H-1)}dxdy = (k+l)^{4H-4}+O((k+l)^{4H-5}). \qquad (5.11)$$

Hence, thanks to (5.9)–(5.11), we observe

$$\begin{aligned}
&\| h_n - h \|^2_{\mathcal{H}^{\odot 2}} = \\
&= n^{2-4H}\sum_{k,l=0}^{n-1} 2H^2(2H-1)^2\left(\int_0^1\int_0^1 (x+y+k+l)^{2(H-1)}dxdy\right)^2 \\
&- 4H^2(2H-1)^2n^{2-4H}\sum_{k,l=0}^{n-1}\int_0^1\left(\int_0^1 (u+v+k+l)^{2H-2}du\right)^2 dv \\
&+ 2H^2(2H-1)^2n^{2-4H}\sum_{k,l=0}^{n-1}\int_0^1\int_0^1 (x+y+k+l)^{4(H-1)}dxdy \\
&\leq Cn^{2-4H}\sum_{k,l=1}^{n-1}(k+l)^{4H-5}
\end{aligned}$$

$$\leq Cn^{3-4H}\sum_{k=1}^{n-1}k^{4H-5}.$$

Here $\sum_{k=1}^{n-1}k^{4H-5}\leq\sum_{k=1}^{\infty}k^{4H-5}<\infty$ since $4H-5<-1$. Thus we obtained $\| h_n-h\|^2_{\mathcal{H}^{\odot 2}}\leq Cn^{3-4H}$ which concludes the proof. □

Proof (*Proof of Theorem* 5.4) We have

$$\frac{n^{2H}}{\sqrt{n}}\sum_{i=0}^{n-1}\left(X_{\frac{i+1}{n}}-X_{\frac{i}{n}}\right)^2\leq n^{2H-3/2}\sup_{0\leq i\leq n-1}\frac{|X_{\frac{i+1}{n}}-X_{\frac{i}{n}}|}{1/n}\sum_{i=0}^{n-1}|X_{\frac{i+1}{n}}-X_{\frac{i}{n}}|.$$

from the properties of the process X

$$\sup_{0\leq i\leq n-1}\frac{|X_{\frac{i+1}{n}}-X_{\frac{i}{n}}|}{1/n}<\infty$$
$$\sum_{i=0}^{n-1}|X_{\frac{i+1}{n}}-X_{\frac{i}{n}}|<\infty.$$

Then

$$\lim_{n\to\infty}\frac{n^{2H}}{\sqrt{n}}\sum_{i=0}^{n-1}\left(X_{\frac{i+1}{n}}-X_{\frac{i}{n}}\right)^2=0\Longleftrightarrow H<3/4.$$

In Addition for $H\in(0,3/4)$ we have as $n\to\infty$:

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left[n^{2H}\left(B^H_{\frac{k+1}{n}}-B^H_{\frac{k}{n}}\right)^2-1\right]\to\mathcal{N}(0,\sigma_H^B)$$

and

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left[n^{2H}\left(S^H_{\frac{k+1}{n}}-S^H_{\frac{k}{n}}\right)^2-1+\alpha_k(H)\right]\to\mathcal{N}(0,\sigma_H^S),$$

where B^H is fBm, S^H is subfBm and $\alpha_k(H)$ is a function defined as follows

$$\alpha_k(H)=H(2H-1)\int_0^1\int_0^1(2k+x+y)^{2(H-1)}dxdy\ \ k\geq 0,$$

such that $\sigma_H^S=\sigma_H^B=\frac{1}{2}\sqrt{\sum_{r\in\mathbb{Z}}\rho(r)}<\infty$. Thus the proof of Theorem 5.4 is done. □

Appendix

Proof (*Proof of Lemma* 5.2) **(i)** We have

$$\alpha(x,x) = (2x+2)^{2H} - 2(2x+1)^{2H} + (2x)^{2H}.$$

Hence by applying Taylor's theorem we observe

$$\alpha(x,x) = 2H(2H-1)(2x)^{2H-2} + O(x^{2H-3})$$

from which the claim follows.
(ii) We show that for any fixed x (resp. y), the function $x \mapsto \alpha(x,y)$ (resp. $y \mapsto \alpha(x,y)$) is decreasing provided that $H \in \left[\frac{3}{4}, 1\right)$. For this, let y be fixed. Then

$$\begin{aligned}\partial_x \alpha(x,y) = 2H \big[&(x+y+2)^{2H-1} - (x+y+1)^{2H-1} \\ &-(x+y+1)^{2H-1} + (x+y)^{2H-1}\big] \\ &= 2H\left[z_y(x+1) - z_y(x)\right],\end{aligned}$$

where $z_y(x) = (x+y+1)^{2H-1} - (x+y)^{2H-1}$. Here

$$z_y'(x) = (2H-1)\left[(x+y+1)^{2H-2} - (x+y)^{2H-2}\right] \le 0,$$

and thus $z_y(x)$ is decreasing. This implies that $\alpha(x,y)$ is also decreasing as a function of x. Furthermore, by changing the roles of x and y we observe that for each fixed x, the function $y \mapsto \alpha(x,y)$ is also decreasing. Consequently, for every $0 \le x \le y$ we have $\alpha(y,y) \le \alpha(x,y) \le \alpha(x,x)$. □

Proof (*Proof of Lemma* 5.3) Using (5.3) and (5.4) we have, for any $H \in (0,1) \setminus \{\frac{1}{2}\}$,

$$\mathbb{E}[V_n^2] = \frac{1}{2}\sum_{r,k=0}^{n-1} \alpha(k,r)^2 = \sum_{r<k}^{n-1} \alpha(k,r)^2 + \frac{1}{2}\sum_{k=0}^{n-1} \alpha(k,k)^2 = I(n) + J(n).$$

The idea is to show that, depending on the value of H, the term $I(n)$ behaves asymptotically correctly while the term $J(n)$ is negligible.
(i) Case $H = \frac{3}{4}$: By Lemma 5.2 item (i) we have

$$J(n) \le C\sum_{k=1}^{n-1} k^{-1} \le C\log n,$$

and hence, as $n \to \infty$, $\frac{J(n)}{n} \to 0$. For the term $I(n)$ we have $\alpha(r,k)^2 = (k+r)^{-1} + O\left((k+r)^{-2}\right)$. Here

$$\frac{1}{n}\sum_{r<k}^{n-1}(k+r)^{-2} \le \frac{C}{n}\sum_{k=1}^{n-1}k^{-1} \le \frac{C\log n}{n} \to 0$$

and consequently the term $O\left((k+r)^{-2}\right)$ does not affect the asymptotic variance. For the term $(k+r)^{-1}$ we substitute $s = k+r$ to get

$$\sum_{r<k}^{n-1}(k+r)^{-1} = \sum_{k=0}^{n-1}\sum_{s=k+1}^{2k-1}s^{-1}.$$

Changing the order of summation we get

$$\sum_{k=0}^{n-1}\sum_{s=k+1}^{2k-1}s^{-1} = \sum_{s=1}^{2n-3}\sum_{\frac{s+1}{2}\le k\le s-1}s^{-1}.$$

Here $\sum_{\frac{s+1}{2}\le k\le s-1}1 = \frac{s}{2}-\frac{1}{2}$ if s is odd and $\sum_{\frac{s+1}{2}\le k\le s-1}1 = \frac{s}{2}$ if s is even. This implies $cn \le \sum_{r<k}^{n-1}(k+r)^{-1} \le Cn$.

(ii) Case $\frac{3}{4} < H < 1$: By Lemma 5.2 item (i) we have $J(n) \le C\sum_{k=1}^{n-1}k^{4H-4} \le Cn^{4H-3}$ since $4H-4 > -1$. Hence, as $n\to\infty$, $\frac{J(n)}{n^{4H-2}} \to 0$. For the term $I(n)$ we use Lemma 5.2 item (ii) to observe that

$$\sum_{r<k}^{n-1}\alpha(k,k)^2 \le \sum_{r<k}^{n-1}\alpha(k,r)^2 \le \sum_{r<k}^{n-1}\alpha(r,r)^2.$$

Here

$$\sum_{r<k}^{n-1}\alpha(k,k)^2 \ge C\sum_{k=1}^{n-1}\left[kk^{4H-4}+O(k^{4H-4})\right].$$

Here $\frac{1}{n^{4H-2}}\sum_{k=1}^{n-1}k^{4H-4} \le Cn^{-1} \to 0$ from which it follows that $\sum_{r<k}^{n-1}\alpha(k,k)^2 \ge cn^{4H-2}$. On the other hand, we have

$$\sum_{r<k}^{n-1}\alpha(r,r)^2 \le C\sum_{r=1}^{n-1}(n-r)r^{4H-4} \le Cn^{4H-2}$$

from which the claim follows. □

References

1. Baxter, G.: A strong limit theorem for Gaussian processes. Proc. Am. Soc. **7**, 522–527 (1997)

2. Begyn, A.: Quadratic variations along irregular subdivisions for Gaussian processes. Electron. J. Probab. **10**(20), 691–717 (2005)
3. Begyn, A.: Asymptotic expansion and central limit theorem for quadratic variations of Gaussian processes. Bernoulli **13**(3), 712–753 (2007)
4. Begyn, A.: Functional limit theorems for generalized quadratic variations of Gaussian processes. Stoch. Process. Appl. **117**, 1848–1869 (2007)
5. Benassi, A., Cohen, S., Istas, J.: Identifying the multifractional function of a Gaussian process. Stat. Probab. Lett. **39**, 337–345 (1998)
6. Benassi, A., Cohen, S., Istas, J., Jaffard, S.: Identification of filtered white noises. Stoch. Probab. Appl. **75**, 31–49 (1998)
7. Bercu, B., Nourdin, I., Taqqu, M.S.: Almost sure central limit theorems on the Wiener space. Stoch. Process. **120**, 1607–1628 (2010)
8. Breton, J.C., Nourdin, I.: Error bounds on the non-normal approximation of Hermite power variations of fractional Brownian motion. Electron. Commun. Probab. **13**, 482–493 (2008)
9. Breuer, P., Major, P.: Central limit theorems for non-linear functionals of Gaussian fields. J. Multivar. Anal. **13**, 425–441 (1983)
10. Coeurjolly, J.: Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. Stat. Inference Stoch. Process. **4**, 199–227 (2001)
11. Coeurjolly, J.: Identification of the multifractional Brownian motion. Bernoulli **11**, 987–1009 (2005)
12. Cohen, S., Guyon, X., Perrin, O., Pontier, M.: Singularity functions for fractional processes: application to fractional Brownian sheet. Annales de l'Institut Henri Poincare **42**(2), 187–205 (2006)
13. De La Vega, W.F.: On almost sure convergence of quadratic Brownian variation. Ann. Probab. **2**, 551–552 (1974)
14. Dobrushin, R.L., Major, P.: Non-central limit theorems for non-linear functionals of Gaussian fields. Z. Wahrsch. verw. Gebiete **50**, 27–52 (1979)
15. Dudley, R.M.: Sample functions of the Gaussian process. Ann. Probab. **1**, 66–103 (1973)
16. Giraitis, L., Surgailis, D.: CLT and other limit theorems for functionals of Gaussian processes. Z. Wahrsch. verw. Gebiete **70**, 191–212 (1985)
17. Gladyshev, E.G.: A new limit theorem for stochastic processes with Gaussian increments. Theory Probab. Appl. **6**(1), 52–61 (1961)
18. Guyon, X., Leon, J.: Convergence en loi des h-variations d'un processus Gaussien stationnaire sur R. Annales de l'Institut Henri Poincare **25**(3), 265–282 (1989)
19. Harnett, D., Nualart, D.: Decomposition and limit theorems for a class of self-similar Gaussian processes (2015). Submitted. arXiv:1508.06641
20. Istas, J., Lang, G.: Quadratic variations and estimation of the local holder index of a Gaussian process. Annal. de l'Institut Henri Poincare **33**(4), 407–436 (1997)
21. Klein, R., Gine, E.: On quadratic variations of processes with Gaussian increments. Ann. Probab. **3**(4), 716–721 (1975)
22. Kubilius, K., Melichov, D.: On the convergence rates of Gladyshev's Hurst index estimator. Nonlinear Anal. Model. Control **15**, 445–450 (2010)
23. Lacaux, C.: Real harmonizable multifractional Lévy motions. Annales de l'Institut Henri Poincare **40**(3), 259–277 (2004)
24. Lei, P., Nualart, D.: A decomposition of the bi-fractional Brownian motion and some applications. Stat. Probab. Lett. **79**, 619–624 (2009)
25. Levental, S., Erickson, R.V.: On almost sure convergence of the quadratic variation of Brownian motion. Stoch. Process. Appl. **106**(2), 317–333 (2003)
26. Malukas, R.: Limit theorems for a quadratic variation of Gaussian processes. Nonlinear Anal. Model. Control **16**(4), 435–452 (2011)
27. Norvaisa, R.: A complement to Gladyshevs theorem. Lith. Math. J. **51**, 26–35 (2011)
28. Nourdin, I., Peccati, G.: Stein's method on Wiener chaos. Probab. Theory Relat. Fields **145**, 1–2 (2009)
29. Nualart, D.: The Malliavin Calculus and Related Topics, 2nd edn. Springer, Berlin (2006)

30. Peccati, G., Taqqu, M.: Wiener Chaos: Moments, Cumulants and Diagrams. A Survey with Computer Implementation. Bocconi University Press, Springer, Berlin (2011)
31. Perrin, O.: Quadratic variation for Gaussian processes and application to time deformation. Stoch. Process. Appl. **40**(3), 293–305 (1999)
32. Ruiz de Chavez, J., Tudor, C.: A decomposition of subfractional Brownian motion. Math. Rep. **61**, 67–74 (2009)
33. Taqqu, M.: Weak convergence to fractional Brownian motion and to the Rosenblatt process. Probab. Theory Relat. Fields **31**(4), 287–302 (1975)
34. Taqqu, M.: Convergence of integrated processes of arbitrary Hermite rank. Z. Wahrsch. Verw. Gebiete **50**, 53–83 (1979)
35. Viitasaari, L.: Sufficient and necessary conditions for limit theorems for quadratic variations of Gaussian sequences (2015). Submitted. arXiv:1502.01370

Chapter 6
Parameter Estimation for Gaussian Processes with Application to the Model with Two Independent Fractional Brownian Motions

Yuliya Mishura, Kostiantyn Ralchenko and Sergiy Shklyar

Abstract The purpose of the article is twofold. Firstly, we review some recent results on the maximum likelihood estimation in the regression model of the form $X_t = \theta G(t) + B_t$, where B is a Gaussian process, $G(t)$ is a known function, and θ is an unknown drift parameter. The estimation techniques for the cases of discrete-time and continuous-time observations are presented. As examples, models with fractional Brownian motion, mixed fractional Brownian motion, and sub-fractional Brownian motion are considered. Secondly, we study in detail the model with two independent fractional Brownian motions and apply the general results mentioned above to this model.

Keywords Discrete observations · Continuous observations · Maximum likelihood estimator · Strong consistency · Fractional Brownian motion · Fredholm integral equation of the first kind

6.1 Introduction

Gaussian processes with drift arise in many applied areas, in particular, in telecommunication and on financial markets. An observed process often can be decomposed as the sum of a useful signal and a random noise, where the last one mentioned is usually modeled by a centered Gaussian process, see, e.g., [19, Ch. VII].

Y. Mishura (✉) · K. Ralchenko · S. Shklyar
Department of Probability Theory, Statistics and Actuarial Mathematics,
Taras Shevchenko National University of Kyiv, 64 Volodymyrska, Kyiv 01601, Ukraine
e-mail: myus@univ.kiev.ua

K. Ralchenko
e-mail: k.ralchenko@gmail.com

S. Shklyar
e-mail: shklyar@univ.kiev.ua

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_6

The simplest example of such model is the process

$$Y_t = \theta t + W_t,$$

where W is a Wiener process. In this case the MLE of the drift parameter θ by observations of Y at points $0 = t_0 \le t_1 \le \cdots \le t_N = T$ is given by

$$\hat{\theta} = \frac{1}{t_N - t_0} \sum_{i=0}^{N-1} \left(Y_{t_{i+1}} - Y_{t_i} \right) = \frac{Y_T - Y_0}{T},$$

and depends on the observations at two points, see e.g. [5]. Models of such type are widely used in finance. For example, Samuelson's model [42] with constant drift parameter μ and known volatility σ has the form

$$\log S_t = \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t,$$

and the MLE of μ equals

$$\hat{\mu} = \frac{\log S_T - \log S_0}{T} + \frac{\sigma^2}{2}.$$

At the same time, the model with Wiener process is not suitable for many processes in natural sciences, computer networks, financial markets, etc., that have long- or short-term dependencies, i.e., the correlations of random noise in these processes decrease slowly with time (long-term dependence) or rapidly with time (short-term dependence). In particular, the models of financial markets demonstrate various kinds of memory (short or long). However, a Wiener process has independent increments, and, therefore, the random noise generated by it is "white", i.e., uncorrelated. The most simple way to overcome this limitation is to use fractional Brownian motion. In some cases even more complicated models are needed. For example, the noise can be modeled by mixed fractional Brownian motion [9], or by the sum of two fractional Brownian motions [28]. Moreover, recently Gaussian processes with non-stationary increments have become popular such as sub-fractional [6], bifractional [14] and multifractional [2, 36, 40] Brownian motions.

In this paper we study rather general model where the noise is represented by a centered Gaussian process $B = \{B_t, t \ge 0\}$ with known covariance function, $B_0 = 0$. We assume that all finite-dimensional distributions of the process $\{B_t,\ t > 0\}$ are multivariate normal distributions with nonsingular covariance matrices. We observe the process X_t with a drift $\theta G(t)$, that is,

$$X_t = \theta G(t) + B_t, \tag{6.1}$$

where $G(t) = \int_0^t g(s)\, ds$, and $g \in L_1[0, t]$ for any $t > 0$.

The paper is devoted to the estimation of the parameter θ by observations of the process X. We consider the MLEs for discrete and continuous schemes of observations. The results presented are based on the recent papers [33, 34]. Note that in [33] the model (6.1) with $G(t) = t$ was considered, and the driving process B was a process with stationary increments. Then in [34] these results were extended to the case of non-linear drift and more general class of driving processes. In the present paper we apply the theoretical results mentioned above to the models with fractional Brownian motion, mixed fractional Brownian motion and sub-fractional Brownian motion.

Similar problems for the model with linear drift driven by fractional Brownian motion were studied in [5, 17, 26, 35]. The mixed Brownian—fractional Brownian model was treated in [7]. In [4, 39] the nonparametric functional estimation of the drift of a Gaussian processes was considered (such estimators for fractional and subfractional Brownian motions were studied in [13, 43] respectively).

In the present paper special attention is given to the model of the form $X_t = \theta t + B_t^{H_1} + B_t^{H_2}$ with two independent fractional Brownian motions B^{H_1} and B^{H_2}. This model was first studied in [28], where a strongly consistent estimator for the unknown drift parameter θ was constructed for $1/2 < H_1 < H_2 < 1$ and $H_2 - H_1 > 1/4$ by continuous-time observations of X. Later, in [30], the strong consistency of this estimator was proved for arbitrary $1/2 < H_1 < H_2 < 1$. The details on this approach are given in Remark 6.2 below. However, the problem of drift parameter estimation by discrete observations in this model was still open. Applying our technique, we obtain the discrete-time estimator of θ and prove its strong consistency for any $H_1, H_2 \in (0, 1)$. Moreover, we also construct the continuous-time estimator and prove the convergence of the discrete-time estimator to the continuous-time one in the case where $H_1 \in (1/2, 3/4]$ and $H_2 \in (H_1, 1)$.

It is worth mentioning that the drift parameter estimation is developed for more general models involving fBm. In particular, the fractional Ornstein–Uhlenbeck process is a popular and well-studied model with fBm. The MLE of the drift parameter for this process was constructed in [20] and further investigated in [3, 44, 47]. Several non-standard estimators for the drift parameter of an ergodic fractional Ornstein–Uhlenbeck process were proposed in [15] and studied in [18]. The corresponding non-ergodic case was treated in [1, 10, 45]. In the papers [8, 11, 12, 16, 48, 49] drift parameter estimators were constructed via discrete observations. More general fractional diffusion models were studied in [21, 27] for continuous-time estimators and in [23, 29, 32] for the case of discrete observations. An estimator of the volatility parameter was constructed in [25]. For Hurst index estimators see, e.g., [22] and references cited therein. Mixed diffusion model including fractional Brownian motion and Wiener process was investigated in [21]. We refer to the paper [31] for a survey of the results on parameter estimation in fractional and mixed diffusion models and to the books [24, 38] for a comprehensive study of this topic.

The paper is organized as follows. In Sect. 6.2 we construct the MLE by discrete-time observations and formulate the conditions for its strong consistency. In Sect. 6.3 we consider the estimator constructed by continuous-time observations and the relations between discrete-time and continuous-time estimators. In Sect. 6.4 these results

are applied to various models mentioned above. In particular, the new approach to parameter estimation in the model with two independent fractional Brownian motions is presented in Sect. 6.4.5. Auxiliary results are proved in Sects. 6.5 and 6.6.

6.2 Construction of Drift Parameter Estimator for Discrete-Time Observations

Let the process X be observed at the points $0 < t_1 < t_2 < \cdots < t_N$. Then the vector of increments

$$\Delta X^{(N)} = (X_{t_1},\ X_{t_2} - X_{t_1},\ \ldots,\ X_{t_N} - X_{t_{N-1}})^\top$$

is a one-to-one function of the observations. We assume in this section that the inequality $G(t_k) \neq 0$ holds at least for one k.

Evidently, vector $\Delta X^{(N)}$ has Gaussian distribution $\mathcal{N}(\theta \Delta G^{(N)}, \Gamma^{(N)})$, where

$$\Delta G^{(N)} = \big(G(t_1),\ G(t_2) - G(t_1),\ \ldots,\ G(t_N) - G(t_{N-1})\big)^\top.$$

Let $\Gamma^{(N)}$ be the covariance matrix of the vector

$$\Delta B^{(N)} = (B_{t_1},\ B_{t_2} - B_{t_1},\ \ldots,\ B_{t_N} - B_{t_{N-1}})^\top.$$

The density of the distribution of $\Delta X^{(N)}$ w. r. t. the Lebesgue measure is

$$\begin{aligned}\mathrm{pdf}_{\Delta X^{(N)}}(x) &= \\ &= \frac{(2\pi)^{-N/2}}{\sqrt{\det \Gamma^{(N)}}} \exp\left\{-\frac{1}{2}\left(x - \theta \Delta G^{(N)}\right)^\top \left(\Gamma^{(N)}\right)^{-1} \left(x - \theta \Delta G^{(N)}\right)\right\}.\end{aligned}$$

Then one can take the density of the distribution of the vector $\Delta X^{(N)}$ for a given θ w. r. t. the density for $\theta = 0$ as a likelihood function:

$$\begin{aligned}L^{(N)}(\theta) &= \\ &= \exp\left\{\theta (\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta X^{(N)} - \frac{\theta^2}{2} (\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta G^{(N)}\right\}.\end{aligned} \tag{6.2}$$

The corresponding MLE equals

$$\hat{\theta}^{(N)} = \frac{\left(\Delta G^{(N)}\right)^\top \left(\Gamma^{(N)}\right)^{-1} \Delta X^{(N)}}{\left(\Delta G^{(N)}\right)^\top \left(\Gamma^{(N)}\right)^{-1} \Delta G^{(N)}}. \tag{6.3}$$

Theorem 6.1 (Properties of the discrete-time MLE [34]) *1. The estimator $\hat{\theta}^{(N)}$ is unbiased and normally distributed:*

$$\hat{\theta}^{(N)} - \theta \simeq \mathcal{N}\left(0, \frac{1}{(\Delta G^{(N)})^{\top}(\Gamma^{(N)})^{-1}\Delta G^{(N)}}\right).$$

2. Assume that

$$\frac{\operatorname{var} B_t}{G^2(t)} \to 0, \quad as\ t \to \infty. \tag{6.4}$$

If $t_N \to \infty$, as $N \to \infty$, then the discrete-time MLE $\hat{\theta}^{(N)}$ converges to θ as $N \to \infty$ almost surely and in $L_2(\Omega)$.

6.3 Construction of Drift Parameter Estimator for Continuous-Time Observations

In this section we suppose that the process X_t is observed on the whole interval $[0, T]$. We investigate MLE for the parameter θ based on these observations.

Let $\langle f,\ g\rangle = \int_0^T f(t)g(t)\,dt$. Assume that the function G and the process B satisfy the following conditions.

(A) There exists a linear self-adjoint operator $\Gamma = \Gamma_T : L_2[0, T] \to L_2[0, T]$ such that

$$\operatorname{cov}(X_s, X_t) = \mathsf{E}\, B_s B_t = \int_0^t \Gamma_T\, 1_{[0,s]}(u)\,du = \langle \Gamma_T\, 1_{[0,s]},\ 1_{[0,t]}\rangle. \tag{6.5}$$

(B) The drift function G is not identically zero, and in its representation $G(t) = \int_0^t g(s)\,ds$ the function $g \in L_2[0, T]$.

(C) There exists a function $h_T \in L_2[0, T]$ such that $g = \Gamma h_T$.

Note that under assumption (A) the covariance between integrals of deterministic functions $f \in L_2[0, T]$ and $g \in L_2[0, T]$ w. r. t. the process B equals

$$\mathsf{E}\int_0^T f(s)\,dB_s \int_0^T g(t)\,dB_t = \langle \Gamma_T f,\ g\rangle.$$

Theorem 6.2 (Likelihood function and continuous-time MLE [34]) *Let T be fixed, assumptions (A)–(C) hold. Then one can choose*

$$L(\theta) = \exp\left\{\theta \int_0^T h_T(s)\,dX_s - \frac{\theta^2}{2}\int_0^T g(s)h_T(s)\,ds\right\} \tag{6.6}$$

as a likelihood function. The MLE equals

$$\hat{\theta}_T = \frac{\int_0^T h_T(s)\,dX_s}{\int_0^T g(s)h_T(s)\,ds}. \tag{6.7}$$

It is unbiased and normally distributed:

$$\hat{\theta}_T - \theta \simeq \mathcal{N}\left(0, \frac{1}{\int_0^T g(s)h_T(s)\,ds}\right).$$

Theorem 6.3 (Consistency of the continuous-time MLE [34]) *Assume that assumptions* (A)–(C) *hold for all* $T > 0$. *If, additionally,*

$$\liminf_{t\to\infty} \frac{\operatorname{var} B_t}{G(t)^2} = 0, \tag{6.8}$$

then the estimator $\hat{\theta}_T$ *converges to* θ *as* $T \to \infty$ *almost surely and in mean square.*

Theorem 6.4 (Relations between discrete and continuous MLEs [34]) *Let the assumptions of Theorem 6.2 hold. Construct the estimator* $\hat{\theta}^{(N)}$ *from (6.3) by observations* $X_{Tk/N}$, $k = 1, \ldots, N$. *Then*

1. *the estimator* $\hat{\theta}^{(N)}$ *converges to* $\hat{\theta}_T$ *in mean square, as* $N \to \infty$,
2. *the estimator* $\hat{\theta}^{(2^n)}$ *converges to* $\hat{\theta}_T$ *almost surely, as* $n \to \infty$.

6.4 Application of Estimators to Models with Various Noises

6.4.1 The Model with Fractional Brownian Motion and Linear Drift

Definition 6.1 The *fractional Brownian motion* $B^H = \left\{B_t^H, t \geq 0\right\}$ with Hurst index $H \in (0, 1)$ is a centered Gaussian process with $B_0 = 0$ and covariance function

$$\mathsf{E}\, B_t^H B_s^H = \tfrac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

Let $H \in (0, 1)$ be fixed. Consider the model

$$X_t = \theta t + B_t^H. \tag{6.9}$$

where X is an observed stochastic process, B^H is an unobserved fractional Brownian motion with Hurst index H, and θ is a parameter of interest. Any finite slice of the stochastic process $\{B_t^H,\ t > 0\}$ has a multivariate normal distribution with nonsingular covariance matrix. Since $\operatorname{var}\left(B_t^H\right) = t^{2H}$, the random process B^H satisfies Theorem 6.1. Hence, we have the following result.

Corollary 6.1 *Under condition* $t_N \to +\infty$ *as* $N \to \infty$, *the estimator* $\hat{\theta}^{(N)}$ *in the model (6.9) is* L_2*-consistent and strongly consistent.*

Remark 6.1 Bertin et al. [5] considered the MLE in the model (6.9) in the discrete scheme of observations, where the trajectory of X was observed at the points $t_k = \frac{k}{N}$, $k = 1, 2, \ldots, N^\alpha$, $\alpha > 1$. Hu et al. [17] investigated the MLE by discrete observations at the points $tk = kh$, $k = 1, 2, \ldots, N$. They considered even more general model of the form $X_t = \theta t + \sigma B_t^H$ with unknown σ. In both papers L_2-consistency and strongly consistency of the MLEs were proved. Note that in Corollary 6.1 both these schemes of observations are allowed, since the only condition $t_N \to \infty$ is required.

Now we consider the case of continuous-time observations and apply the results of Sect. 6.3 to the model (6.9). Let $H \in (\frac{1}{2}, 1)$. Denote by Γ_H the corresponding operator Γ for the model (6.9). Then

$$(\Gamma_H f)(t) = H(2H-1) \int_0^T \frac{f(s)}{|t-s|^{2-2H}}\, ds. \tag{6.10}$$

For the function

$$h_T(s) = C_H s^{1/2-H}(T-s)^{1/2-H},$$

$C_H = \left(H(2H-1)\mathrm{B}\left(H-\frac{1}{2}, \frac{3}{2}-H\right)\right)^{-1}$, we have that

$$\Gamma_H h_T = 1_{[0,T]}, \tag{6.11}$$

see [35]. The MLE is given by

$$\hat{\theta}_T = \frac{T^{2H-2}}{\mathrm{B}(3/2-H,\ 3/2-H)} \int_0^T s^{1/2-H}(T-s)^{1/2-H}\, dX_s.$$

This estimator was studied in [26, 35], see also [33, Example 3.11].

Corollary 6.2 *Let $H \in (\frac{1}{2}, 1)$. The conditions of Theorems 6.2, 6.3 and 6.4, are satisfied. The estimator $\hat{\theta}_T$ is L_2-consistent and strongly consistent. For fixed T, it can be approximated by discrete-sample estimator in mean-square sense.*

6.4.2 Model with Fractional Brownian Motion and Power Drift

Now we generalize the model (6.9) for the case of the non-linear drift function $G(t) = t^{\alpha+1}$. Let $0 < H < 1$ and $\alpha > -1$. Consider the process

$$X_t = \theta t^{\alpha+1} + B_t^H \tag{6.12}$$

This is a particular case of model (6.1), with $g(t) = (\alpha+1)t^\alpha$.

Now verify the conditions of the theorems. The condition (6.4) holds true if and only if $\alpha > H - 1$.

Corollary 6.3 *If $\alpha > H - 1$, the model (6.12) satisfies the conditions of Theorem 6.1. The estimator $\hat{\theta}^{(N)}$ in the model (6.12) is L_2-consistent and strongly consistent (provided that $\lim_{N\to\infty} t_N = +\infty$).*

The condition (B): $g \in L_2[0, T]$ holds true if and only if $\alpha > -\frac{1}{2}$. The integral equation $\Gamma h = g$ is rewritten as

$$\int_0^T \frac{h(s)\,ds}{|t-s|^{2H-2}} = \frac{\alpha+1}{(2H-1)H} t^{\alpha}.$$

If $\alpha > 2H - 2$, then the solution is

$$h(t) = \text{const} \cdot \left(\frac{T^{\alpha}}{t^{H-\frac{1}{2}}(T-t)^{H-\frac{1}{2}}} - \alpha t^{\alpha+1-2H} W\left(\tfrac{T}{t},\ \alpha,\ H-\tfrac{1}{2}\right) \right), \tag{6.13}$$

where $W\left(\frac{T}{t},\ \alpha,\ H-\frac{1}{2}\right) = \int_0^{\frac{T}{t}-1} (v+1)^{\alpha-1} v^{\frac{1}{2}-H}\,dv$. The asymptotic behaviour of the function $W\left(\frac{T}{t},\ \alpha,\ H-\frac{1}{2}\right)$ as $t \to 0+$ is

$$W\left(\tfrac{T}{t},\ \alpha,\ H-\tfrac{1}{2}\right) \sim \begin{cases} \mathrm{B}\left(\frac{3}{2}-H,\ H-\frac{1}{2}-\alpha\right) & \text{if } \alpha < H-\frac{1}{2}, \\ \ln(T/t) & \text{if } \alpha = H-\frac{1}{2}, \\ \frac{2}{2\alpha+1-2H} \frac{T^{\alpha-H+\frac{1}{2}}}{t^{\alpha-H+\frac{1}{2}}} & \text{if } \alpha > H-\frac{1}{2}. \end{cases}$$

Therefore, the function $h(t)$ defined in (6.13) is square integrable if $\alpha + 1 - 2H - \max\left(0,\ \alpha - H + \frac{1}{2}\right) > -\frac{1}{2}$, which holds if $\alpha > 2H - \frac{3}{2}$. Note that if $\alpha > 2H - \frac{3}{2}$, then the following inequalities hold true: $\alpha > 2H - 2$ (whence h defined in (6.13) is indeed a solution to the integral equation $\Gamma h = g$), $\alpha > H - 1$ (whence conditions (6.4) and so (6.8) are satisfied), and $\alpha > -\frac{1}{2}$ (whence condition (B) is satisfied).

Corollary 6.4 *If $\alpha > 2H - \frac{3}{2}$, the conditions of Theorems 6.2, 6.3 and 6.4, are satisfied. The estimator $\hat{\theta}_T$ is L_2-consistent and strongly consistent. For fixed T, it can be approximated by discrete-sample estimator in mean-square sense.*

6.4.3 The Model with Brownian and Fractional Brownian Motion

Consider the following model:

$$X_t = \theta t + W_t + B_t^H, \tag{6.14}$$

where W is a standard Wiener process, B^H is a fractional Brownian motion with Hurst index H, and random processes W and B^H are independent. The corresponding operator Γ is $\Gamma = I + \Gamma_H$, where Γ_H is defined by (6.10). The operator Γ_H is self-adjoint and positive semi-definite. Hence, the operator Γ is invertible. Thus Assumption (C) holds true.

In other words, the problem is reduced to the solving of the following Fredholm integral equation of the second kind

$$h_T(u) + H_2(2H_2 - 1) \int_0^T h_T(s) \, |s-u|^{2H_2-2} \, ds = 1, \quad u \in [0, T]. \tag{6.15}$$

This approach to the drift parameter estimation in the model with mixed fractional Brownian motion was first developed in [7].

Note also that the function $h_T = \Gamma_T^{-1} \mathbf{1}_{[0,\mathrm{T}]}$ can be evaluated iteratively

$$h_T = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2} \left\| \Gamma_T^H \right\| I - \Gamma_T^H\right)^k \mathbf{1}_{[0,\mathrm{T}]}}{\left(1 + \frac{1}{2} \left\| \Gamma_T^H \right\|\right)^{k+1}}. \tag{6.16}$$

6.4.4 Model with Subfractional Brownian Motion

Definition 6.2 The *subfractional Brownian motion* $\widetilde{B}^H = \left\{\widetilde{B}_t^H, t \geq 0\right\}$ with Hurst parameter $H \in (0, 1)$ is a centered Gaussian random process with covariance function

$$\operatorname{cov}\left(\widetilde{B}_s^H, \widetilde{B}_t^H\right) = \frac{2\,|t|^{2H} + 2\,|s|^{2H} - |t-s|^{2H} - |t+s|^{2H}}{2}. \tag{6.17}$$

We refer to [6, 46] for properties of this process. Obviously, neither $\widetilde{B}^H$, nor its increments are stationary. If $\{B_t^H, \ t \in \mathbb{R}\}$ is a fractional Brownian motion, then the random process $\frac{B_t^H + B_{-t}^H}{\sqrt{2}}$ is a subfractional Brownian motion. Evidently, mixed derivative of the covariance function (6.17) equals

$$\begin{aligned} K_H(s,t) &:= \frac{\partial^2 \operatorname{cov}\left(\widetilde{B}_s^H, \widetilde{B}_t^H\right)}{\partial t \, \partial s} = \\ &= H\,(2H-1)\left(|t-s|^{2H-2} - |t+s|^{2H-2}\right). \end{aligned} \tag{6.18}$$

If $H \in (\frac{1}{2}, 1)$, then the operator $\Gamma = \widetilde{\Gamma}_H$ that satisfies (6.5) for $\widetilde{B}^H$ equals

$$\widetilde{\Gamma}_H f(t) = \int_0^T K_H(s,t) f(s) \, ds. \tag{6.19}$$

Consider the model (6.1) for $G(t) = t$ and $B = \widetilde{B}^H$:

$$X_t = \theta t + \widetilde{B}^H_t. \tag{6.20}$$

Let us construct the estimators $\hat{\theta}^{(N)}$ and $\hat{\theta}_T$ from (6.3) and (6.7) respectively and establish their properties. In particular, Proposition 6.1 allows to define finite-sample estimator $\hat{\theta}^{(N)}$.

Proposition 6.1 *The linear equation $\widetilde{\Gamma}_H f = 0$ has only trivial solution in $L_2[0, T]$. As a consequence, the finite slice $\left(\widetilde{B}^H_{t_1}, \dots, \widetilde{B}^H_{t_N}\right)$ with $0 < t_1 < \cdots < t_N$ has a multivariate normal distribution with nonsingular covariance matrix.*

Since $\operatorname{var}\left(\widetilde{B}^H_t\right) = \left(2 - 2^{2H-1}\right) t^{2H}$, the random process $\widetilde{B}^H$ satisfies Theorem 6.1. Hence, we have the following result.

Corollary 6.5 *Under condition $t_N \to +\infty$ as $N \to \infty$, the estimator $\hat{\theta}^{(N)}$ in the model (6.20) is L_2-consistent and strongly consistent.*

In order to define the continuous-time MLE (6.7), we have to solve an integral equation. The following statement guarantees the existence of a solution.

Proposition 6.2 *If $\frac{1}{2} < H < \frac{3}{4}$, then the integral equation $\widetilde{\Gamma}_H h = \mathbf{1}_{[0,T]}$, that is*

$$\int_0^T K_H(s,t)h(s)\,ds = 1 \quad \text{for almost all } t \in (0,T) \tag{6.21}$$

has a unique solution $h \in L_2[0, T]$.

Corollary 6.6 *If $\frac{1}{2} < H < \frac{3}{4}$, then the random process $\widetilde{B}^H$ satisfies Theorems 6.2, 6.3, and 6.4. As the result, $L(\theta)$ defined in (6.6) is the likelihood function in the model (6.20), and $\hat{\theta}_T$ defined in (6.7) is the MLE. The estimator is L_2-consistent and strongly consistent. For fixed T, it can be approximated by discrete-sample estimator in mean-square sense.*

6.4.5 The Model with Two Independent Fractional Brownian Motions

Consider the following model:

$$X_t = \theta t + B^{H_1}_t + B^{H_2}_t, \tag{6.22}$$

where B^{H_1} and B^{H_2} are two independent fractional Brownian motion with Hurst indices $H_1, H_2 \in (\frac{1}{2}, 1)$. Obviously, the condition (6.4) is satisfied:

$$\frac{\operatorname{var}\left(B^{H_1}_t + B^{H_2}_t\right)}{t^2} = \frac{t^{2H_1} + t^{2H_2}}{t^2} \to 0, \quad t \to \infty.$$

Theorem 6.5 *Under condition $t_N \to +\infty$ as $N \to \infty$, the estimator $\hat{\theta}^{(N)}$ in the model (6.22) is L_2-consistent and strongly consistent.*

Evidently, the corresponding operator Γ for the model (6.22) equals $\Gamma_{H_1} + \Gamma_{H_2}$, where Γ_H is defined by (6.10). Therefore, in order to verify the assumptions of Theorem 6.2 we need to show that there exists a function h_T such that $\left(\Gamma_{H_1} + \Gamma_{H_2}\right) h_T = \mathbf{1}_{[0,T]}$. This is equivalent to $\left(I + \Gamma_{H_1}^{-1}\Gamma_{H_2}\right) h_T = \Gamma_{H_1}^{-1}\mathbf{1}_{[0,T]}$, since the operator Γ_{H_1} is injective and its range contains $\mathbf{1}_{[0,T]}$, see (6.11) and Theorem 6.7. Hence, it suffices to prove that the operator $I + \Gamma_{H_1}^{-1}\Gamma_{H_2}$ is invertible. This is done in Theorem 6.8 in Sect. 6.6 for $H_1 \in (1/2, 3/4]$, $H_2 \in (H_1, 1)$. Thus, in this case the assumptions of Theorem 6.2 hold with

$$h_T = \left(I + \Gamma_{H_1}^{-1}\Gamma_{H_2}\right)^{-1} \Gamma_{H_1}^{-1}\mathbf{1}_{[0,T]}\,.$$

Therefore, we have the following result for the estimator

$$\hat{\theta}_T = \frac{\int_0^T h_T(s)\,dX_s}{\int_0^T h_T(s)\,ds}.$$

Theorem 6.6 *If $H_1 \in (1/2, 3/4]$ and $H_2 \in (H_1, 1)$, then the random process $B^{H_1} + B^{H_2}$ satisfies Theorems 6.2, 6.3, and 6.4. As the result, $L(\theta)$ defined in (6.6) is the likelihood function in the model (6.22), and $\hat{\theta}_T$ is the maximum likelihood estimator. The estimator is L_2-consistent and strongly consistent. For fixed T, it can be approximated by discrete-sample estimator in mean-square sense.*

Remark 6.2 Another approach to the drift parameter estimation in the model with two fractional Brownian motions was proposed in [28] and developed in [30]. It is based on the solving of the following Fredholm integral equation of the second kind

$$\begin{aligned}(2-2H_1)\tilde{h}_T(u)u^{1-2H_1} + \int_0^T \tilde{h}_T(s)k(s,u)\,ds = \\ = (2-2H_1)u^{1-2H_1}, \quad u \in (0,T],\end{aligned} \tag{6.23}$$

where

$$k(s,u) = \int_0^{s\wedge u} \partial_s K_{H_1,H_2}(s,v)\partial_u K_{H_1,H_2}(u,v)\,dv,$$

$$K_{H_1,H_2}(t,s) = c_{H_1}\beta_{H_2}s^{1/2-H_2}\int_s^t (t-u)^{1/2-H_1}u^{H_2-H_1}(u-s)^{H_2-3/2}du,$$

$$c_{H_1} = \left(\frac{\Gamma(3-2H_1)}{2H_1\Gamma(\frac{3}{2}-H_1)^3\Gamma(H_1+\frac{1}{2})}\right)^{\frac{1}{2}},$$

$$\beta_{H_2} = \left(\frac{2H_2 \left(H_2 - \frac{1}{2}\right)^2 \Gamma(\frac{3}{2} - H_2)}{\Gamma(H_2 + \frac{1}{2})\Gamma(2 - 2H_2)} \right)^{\frac{1}{2}}.$$

Then for $1/2 \le H_1 < H_2 < 1$ the estimator is defined as

$$\hat{\theta}(T) = \frac{N(T)}{\delta_{H_1} \langle N \rangle(T)},$$

where $\delta_{H_1} = c_{H_1} \mathrm{B}\left(\frac{3}{2} - H_1, \frac{3}{2} - H_1\right)$, $N(t)$ is a square integrable Gaussian martingale,

$$N(T) = \int_0^T \tilde{h}_T(t)\, dX(t),$$

$\tilde{h}_T(t)$ is a unique solution to (6.23) and

$$\langle N \rangle(T) = (2 - 2H_1) \int_0^T \tilde{h}_T(t) t^{1-2H_1}\, dt.$$

This estimator is also unbiased, normal and strongly consistent. The details of this method can be found also in [24, Sec. 5.5].

6.5 Integral Equation with Power Kernel

Theorem 6.7 *Let $0 < p < 1$ and $b > 0$.*

1. *If $y \in L_1[0, b]$ is a solution to integral equation*

$$\int_0^b \frac{y(s)\, ds}{|t - s|^p} = f(t) \quad \text{for almost all } t \in (0, b), \tag{6.24}$$

then $y(x)$ satisfies

$$y(x) = \frac{\Gamma(p) \cos \frac{\pi p}{2}}{\pi x^{(1-p)/2}} \mathcal{D}_{b-}^{(1-p)/2} \left(x^{1-p} \mathcal{D}_{0+}^{(1-p)/2} \left(\frac{f(x)}{x^{(1-p)/2}} \right) \right) \tag{6.25}$$

almost everywhere on $[0, b]$, where $\mathcal{D}_{a+}^{\alpha}$ and $\mathcal{D}_{b-}^{\alpha}$ are the Riemann–Liouville fractional derivatives, that is

$$\mathcal{D}_{a+}^{\alpha} f(x) = \frac{1}{\Gamma(1 - \alpha)} \frac{d}{dx} \left(\int_a^x \frac{f(t)}{(x - t)^{\alpha}}\, dt \right),$$
$$\mathcal{D}_{b-}^{\alpha} f(x) = \frac{-1}{\Gamma(1 - \alpha)} \frac{d}{dx} \left(\int_x^b \frac{f(t)}{(t - x)^{\alpha}}\, dt \right).$$

2. *If $y_1 \in L_1[0,b]$ and $y_2 \in L_1[0,b]$ are two solutions to integral equation (6.24), then $y_1(x) = y_2(x)$ almost everywhere on $[0,b]$.*

3. *If $y \in L_1[0,b]$ satisfies (6.25) almost everywhere on $[0,b]$ and the fractional derivatives are solutions to respective Abel integral equations, that is*

$$\frac{1}{\Gamma\left(\frac{1-p}{2}\right)} \int_0^t \frac{\mathcal{D}_{0+}^{(1-p)/2}(f(x)x^{(p-1)/2})}{(t-x)^{(p+1)/2}}\,dx = \frac{f(t)}{t^{(1-p)/2}}, \tag{6.26}$$

for almost all $t \in (0,b)$ and

$$\frac{1}{\Gamma\left(\frac{1-p}{2}\right)} \int_x^b \frac{\pi y(s) s^{(1-p)/2}}{\Gamma(p)\cos\frac{\pi p}{2}} \frac{ds}{(s-x)^{(p+1)/2}} = \tag{6.27}$$

$$= x^{1-p}\mathcal{D}_{0+}^{(1-p)/2}\left(\frac{f(x)}{x^{(1-p)/2}}\right)$$

for almost all $x \in (0,b)$, then $y(s)$ is a solution to integral equation (6.24).

Proof Firstly, transform the left-hand side of (6.24). By [35, Lemma 2.2(i)], for $0 < s < t$

$$\int_0^s \frac{d\tau}{(t-\tau)^{(p+1)/2}(s-\tau)^{(p+1)/2}\tau^{1-p}} = \frac{\mathrm{B}\left(p,\ \frac{1-p}{2}\right)}{s^{(1-p)/2}t^{(1-p)/2}(t-s)^p}.$$

Hence, for $s > 0$, $t > 0$, $s \neq t$

$$\int_0^{\min(s,t)} \frac{d\tau}{(t-\tau)^{(p+1)/2}(s-\tau)^{(p+1)/2}\tau^{1-p}} = \frac{\mathrm{B}\left(p,\ \frac{1-p}{2}\right)}{s^{(1-p)/2}t^{(1-p)/2}|t-s|^p}.$$

Hence

$$\int_0^b \frac{y(s)\,ds}{|t-s|^p} =$$

$$= \int_0^b \frac{s^{(1-p)/2}t^{(1-p)/2}y(s)}{\mathrm{B}\left(p,\ \frac{1-p}{2}\right)} \int_0^{\min(s,t)} \frac{d\tau}{(t-\tau)^{(p+1)/2}(s-\tau)^{(p+1)/2}\tau^{1-p}}\,ds.$$

Change the order of integration, noting that $\{(s,\tau) : 0 < s < b,\ 0 < \tau < \min(s,t)\} = \{(s,\tau) : 0 < \tau < t,\ \tau < s < b\}$ for $0 < t < b$:

$$\int_0^b \frac{y(s)\,ds}{|t-s|^p} = \tag{6.28}$$
$$= \frac{t^{(1-p)/2}}{\mathrm{B}\left(p,\ \frac{1-p}{2}\right)} \int_0^t \frac{1}{(t-\tau)^{(p+1)/2}\tau^{1-p}} \int_\tau^b \frac{s^{(1-p)/2} y(s)\,ds}{(s-\tau)^{(p+1)/2}}\,d\tau.$$

The right-hand side of (6.28) can be rewritten with fractional integration:

$$\int_0^b \frac{y(s)\,ds}{|x-s|^p} = \frac{\Gamma\left(\frac{1-p}{2}\right)^2}{\mathrm{B}\left(p,\ \frac{1-p}{2}\right)} x^{(1-p)/2} I_{0+}^{(1-p)/2} \left(\frac{1}{x^{1-p}} I_{b-}^{(1-p)/2}\left(x^{(1-p)/2} y(x)\right)\right)$$

for $0 < x < b$, where I_{a+}^{α} and I_{b-}^{α} are fractional integrals

$$I_{a+}^{\alpha} f(x) = \frac{1}{\Gamma(\alpha)} \int_a^x \frac{f(t)}{(x-t)^{(1-\alpha)}}\,dt$$

and

$$I_{b-}^{\alpha} f(x) = \frac{1}{\Gamma(\alpha)} \int_x^b \frac{f(t)}{(t-x)^{(1-\alpha)}}\,dt.$$

The constant coefficient can be simplified:

$$\frac{\Gamma\left(\frac{1-p}{2}\right)^2}{\mathrm{B}\left(p,\ \frac{1-p}{2}\right)} = \frac{\Gamma\left(\frac{1-p}{2}\right)\Gamma\left(\frac{p+1}{2}\right)}{\Gamma(p)} = \frac{\pi}{\Gamma(p)\cos\left(\frac{\pi p}{2}\right)}.$$

Thus integral equation (6.24) can be rewritten with use of fractional integrals:

$$\frac{\pi}{\Gamma(p)\cos\left(\frac{\pi p}{2}\right)} x^{(1-p)/2} I_{0+}^{(1-p)/2} \left(\frac{1}{x^{1-p}} I_{b-}^{(1-p)/2}\left(x^{(1-p)/2} y(x)\right)\right) = f(x) \tag{6.29}$$

for almost all $x \in (0, b)$.

Whenever $y \in L_1[0, b]$, the function $x^{(1-p)/2} y(x)$ is obviously integrable on $[0, b]$. Now prove that the function $x^{p-1} I_{b-}^{(1-p)/2}\left(x^{(1-p)/2} y(x)\right)$ is also integrable on $[0, b]$. Indeed,

$$\left|I_{b-}^{(1-p)/2}\left(x^{(1-p)/2}y(x)\right)\right| \le \frac{1}{\Gamma\left(\frac{1-p}{2}\right)}\int_x^b \frac{t^{(1-p)/2}\,|y(t)|}{(t-x)^{(p+1)/2}}\,dt,$$

$$\begin{aligned}\int_0^b \left|\frac{I_{b-}^{(1-p)/2}\left(x^{(1-p)/2}y(x)\right)}{x^{1-p}}\right| dx &\le \frac{1}{\Gamma\left(\frac{1-p}{2}\right)}\int_0^b \frac{1}{x^{1-p}}\int_x^b \frac{t^{(1-p)/2}\,|y(t)|\,dt}{(t-x)^{(p+1)/2}}\,dx\\ &= \frac{1}{\Gamma\left(\frac{1-p}{2}\right)}\int_0^b t^{(1-p)/2}\,|y(t)|\int_0^t \frac{dx}{x^{1-p}(t-x)^{(p+1)/2}}\,dt\\ &= \frac{\mathrm{B}\left(p,\ \frac{1-p}{2}\right)}{\Gamma\left(\frac{1-p}{2}\right)}\int_0^b |y(t)|\,dt < \infty,\end{aligned}$$

and the integrability is proved.

Due to [41, Theorem 2.1], the Abel integral equation $f(x) = I_{a+}^{\alpha}\phi(x), x \in (a,b)$, may have not more that one solution $\phi(x)$ within $L_1[a,b]$. If the equation has such a solution, then the solution $\phi(x)$ is equal to $\mathcal{D}_{a+}^{\alpha}f(x)$. Similarly, the Abel integral equation $f(x) = I_{b-}^{\alpha}\phi(x)$ may have not more that one solution $\phi(x) \in L_1[a,b]$, and if it exists, $\phi = \mathcal{D}_{b-}^{\alpha}f$.

Therefore, if $y \in L_1[0,b]$ is a solution to integral equation, then is also satisfies (6.29), so

$$\begin{aligned}I_{0+}^{(1-p)/2}\left(\frac{1}{x^{1-p}}I_{b-}^{(1-p)/2}\left(\frac{\pi}{\Gamma(p)\cos\left(\frac{\pi p}{2}\right)}x^{(1-p)/2}y(x)\right)\right) &= \frac{f(x)}{x^{(1-p)/2}}, \qquad (6.30)\\ \frac{1}{x^{1-p}}I_{b-}^{(1-p)/2}\left(\frac{\pi x^{(1-p)/2}y(x)}{\Gamma(p)\cos\left(\frac{\pi p}{2}\right)}\right) &= \mathcal{D}_{0+}^{(1-p)/2}\left(\frac{f(x)}{x^{(1-p)/2}}\right),\\ \frac{\pi x^{(1-p)/2}y(x)}{\Gamma(p)\cos\left(\frac{\pi p}{2}\right)} &= \mathcal{D}_{b-}^{(1-p)/2}\left(x^{1-p}\mathcal{D}_{0+}^{(1-p)/2}\left(\frac{f(x)}{x^{(1-p)/2}}\right)\right)\end{aligned}$$

for almost all $x \in (0,b)$. Thus $y(x)$ satisfies (6.25). Statement 1 of Theorem 6.7 is proved, and statement 2 follows from statement 1.

From Eqs. (6.26) and (6.27), which can be rewritten with fractional integration operator,

$$\begin{aligned}I_{0+}^{(1-p)/2}\mathcal{D}_{0+}^{(1-p)/2}\left(f(x)x^{(1-p)/2}\right) &= \frac{f(x)}{t^{(1-p)/2}},\\ I_{b-}^{(1-p)/2}\left(\frac{\pi y(x)x^{(1-p)/2}}{\Gamma(p)\cos\frac{\pi p}{2}}\right) &= x^{1-p}\mathcal{D}_{0+}^{(1-p)/2}\left(\frac{f(x)}{x^{(1-p)/2}}\right)\end{aligned}$$

(6.30) follows, and (6.30) is equivalent to (6.24). Thus statement 3 of Theorem 6.7 holds true. □

Remark 6.3 The integral equation (6.24) was solved explicitly in [26, Lemma 3] under the assumption $f \in C([0,b])$. Here we solve this equation in $L_1[0,b]$ and prove the uniqueness of a solution in this space. Note also that the formula for solution in the handbook [37, formula 3.1.30] is incorrect (it is derived from the incorrect formula 3.1.32 of the same book, where an operator of differentiation is missing; this error comes from the book [50]).

6.6 Boundedness and Invertibility of Operators

This section is devoted to the proof of the following result, which plays the key role in the proof of the strong consistency of the MLE for the model with two independent fractional Brownian motions.

Theorem 6.8 *Let $H_1 \in \left(\frac{1}{2}, \frac{3}{4}\right]$, $H_2 \in (H_1, 1)$, and Γ_H be the operator defined by (6.10). Then $\Gamma_{H_1}^{-1}\Gamma_{H_2} \colon L_2[0,T] \to L_2[0,T]$ is a compact linear operator defined on the entire space $L_2[0,T]$, and the operator $I + \Gamma_{H_1}^{-1}\Gamma_{H_2}$ is invertible.*

The proof consists of several steps.

6.6.1 Convolution Operator

If $\phi \in L_1[-T,T]$, then the following convolution operator

$$Lf(x) = \int_0^T \phi(t-s)f(s)\,ds \tag{6.31}$$

is a linear continuous operator $L_2[0,T] \to L_2[0,T]$, and

$$\|L\| \le \int_{-T}^{T} |\phi(t)|\,dt. \tag{6.32}$$

Moreover, L is a compact operator.

The adjoint operator of the operator (6.31) is

$$L^* f(x) = \int_0^T \phi(s-t)f(s)\,ds.$$

If the function ϕ is even, then the linear operator L is self-adjoint.

Let us consider the following convolution operators.

Definition 6.3 For $\alpha > 0$, the Riemann–Liouville operators of fractional integration are defined as

$$I_{0+}^{\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t \frac{f(s)\, ds}{(t-s)^{1-\alpha}},$$
$$I_{T-}^{\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_t^T \frac{f(s)\, ds}{(t-s)^{1-\alpha}}.$$

The operators I_{0+}^{α} and I_{T-}^{α} are mutually adjoint. Their norm can be bounded as follows

$$\|I_{T-}^{\alpha}\| = \|I_{0+}^{\alpha}\| \le \frac{1}{\Gamma(\alpha)} \int_0^T \frac{ds}{s^{1-\alpha}} = \frac{T^{\alpha}}{\Gamma(\alpha+1)}. \tag{6.33}$$

Let $\frac{1}{2} < H < 1$ and Γ_H be the operator defined by (6.10). Then

$$\Gamma_H = H\Gamma(2H)\left(I_{0+}^{2H-1} + I_{T-}^{2H-1}\right). \tag{6.34}$$

The linear operators I_{0+}^{α}, I_{T-}^{α} for $\alpha > 0$, and Γ_H for $\frac{1}{2} < H < 1$ are injective.

6.6.2 Semigroup Property of the Operator of Fractional Integration

Theorem 6.9 *For $\alpha > 0$ and $\beta > 0$ the following equalities hold*

$$I_{0+}^{\alpha} I_{0+}^{\beta} = I_{0+}^{\alpha+\beta},$$
$$I_{T-}^{\alpha} I_{T-}^{\beta} = I_{T-}^{\alpha+\beta}.$$

This theorem is a particular case of [41, Theorem 2.5].

Proposition 6.3 *For $0 < \alpha \le \frac{1}{2}$ and $f \in L_2[0, T]$,*

$$\langle I_{0+}^{\alpha} f,\ I_{T-}^{\alpha} f \rangle \ge 0.$$

Equality is achieved if and only if

- $f = 0$ *almost everywhere on* $[0, T]$ *for* $0 < \alpha < \frac{1}{2}$;
- $\int_0^T f(t)\, dt = 0$ *for* $\alpha = \frac{1}{2}$.

Proof Since the operators I_{0+}^{α} and I_{T-}^{α} are mutually adjoint, by semigroup property, we have that

$$\langle I_{0+}^{\alpha} f,\ I_{T-}^{\alpha} f \rangle = \langle I_{0+}^{\alpha} I_{0+}^{\alpha} f,\ f \rangle = \left\langle I_{0+}^{2\alpha} f,\ f \right\rangle,$$
$$\langle I_{0+}^{\alpha} f,\ I_{T-}^{\alpha} f \rangle = \langle f,\ I_{T-}^{\alpha} I_{T-}^{\alpha} f \rangle = \left\langle f,\ I_{T-}^{2\alpha} f \right\rangle.$$

Adding these equalities, we obtain

$$\langle I^{\alpha}_{0+} f,\ I^{\alpha}_{T-} f\rangle = \frac{1}{2}\left\langle I^{2\alpha}_{0+} f + I^{2\alpha}_{T-} f,\ f\right\rangle. \tag{6.35}$$

If $0 < \alpha < \frac{1}{2}$, then

$$\begin{aligned}\langle I^{\alpha}_{0+} f,\ I^{\alpha}_{T-} f\rangle &= \frac{1}{2H\Gamma(2H)}\,\langle \Gamma_H f,\ f\rangle = \\ &= \frac{1}{2H\Gamma(2H)}\,\mathsf{E}\left(\int_0^T f(t)\,dB^H_t\right)^2 \geq 0.\end{aligned}$$

where $H = \alpha + \frac{1}{2}$, $\frac{1}{2} < H < 1$, and B^H_t is a fractional Brownian motion.

Let us consider the case $\alpha = \frac{1}{2}$. Since

$$\begin{aligned} I^1_{0+} f(t) + I^1_{T-} f(t) &= \int_0^t f(s)\,ds + \int_t^T f(s)\,ds = \int_0^T f(s)\,ds,\\ I^1_{0+} f + I^1_{T-} f &= \int_0^T f(s)\,ds\ \mathsf{1}_{[0,T]},\\ \left\langle I^1_{0+} f + I^1_{T-} f,\ f\right\rangle &= \int_0^T f(s)\,ds\ \langle \mathsf{1}_{[0,T]},\ f\rangle = \left(\int_0^T f(s)\,ds\right)^2, \end{aligned}$$

we see from (6.35) that

$$\left\langle I^{1/2}_{0+} f,\ I^{1/2}_{T-} f\right\rangle = \frac{1}{2}\left(\int_0^T f(s)\,ds\right)^2 \geq 0. \tag{6.36}$$

Conditions for the equality $\langle I^{\alpha}_{0+} f,\ I^{\alpha}_{T-} f\rangle = 0$ can be easily found by analyzing the proof. Indeed, if $0 < \alpha < \frac{1}{2}$ and $H{=}\alpha{+}\frac{1}{2}$, then Γ_H is a self-adjoint positive compact operator whose eigenvalues are all positive. Then

$$2H\Gamma(2H)\langle I^{\alpha}_{0+} f,\ I^{\alpha}_{T-} f\rangle = \langle \Gamma_H f,\ f\rangle = \|\Gamma^{1/2}_H f\|^2.$$

In this case, the equality $\langle I^{\alpha}_{0+} f,\ I^{\alpha}_{T-} f\rangle = 0$ holds true if and only if $f = 0$ almost everywhere on $[0, T]$. If $\alpha = \frac{1}{2}$, then the condition for the equality follows from (6.36). □

Proposition 6.4 *For* $0 < \alpha \leq \frac{1}{2}$ *and* $f \in L_2[0, T]$,

$$\|I^{\alpha}_{0+} f\| + \|I^{\alpha}_{T-} f\| \leq \sqrt{2}\,\|I^{\alpha}_{0+} f + I^{\alpha}_{T-} f\|. \tag{6.37}$$

Consequently, for $\frac{1}{2} < H \le \frac{3}{4}$

$$\left\| I_{0+}^{2H-1} f \right\| + \left\| I_{T-}^{2H-1} f \right\| \le \frac{\sqrt{2}}{H\Gamma(2H)} \| \Gamma_H f \|.$$

Proof Taking into account Proposition 6.3, we get

$$\begin{aligned}(\| I_{0+}^{\alpha} f \| + \| I_{T-}^{\alpha} f \|)^2 &\le 2 \| I_{0+}^{\alpha} f \|^2 + 2 \| I_{T-}^{\alpha} f \|^2 \\ &\le 2 \| I_{0+}^{\alpha} f \|^2 + 4 \langle I_{0+}^{\alpha} f,\ I_{T-}^{\alpha} f \rangle + 2 \| I_{T-}^{\alpha} f \|^2 \\ &= 2 \| I_{0+}^{\alpha} f + I_{T-}^{\alpha} f \|^2,\end{aligned}$$

whence the inequality (6.37) follows.

The second statement is obtained by the representation (6.34). □

6.6.3 Transposition of Operators

Lemma 6.1 *Let A be a linear continuous operator on $L_2[0, T]$, and B be an injective self-adjoint compact linear operator on $L_2[0, T]$. If the linear operator $A^* B^{-1}$ is bounded, that is $\| A^* B^{-1} \| = K < \infty$, then the linear operator $B^{-1} A$ is defined on the entire space $L_2[0, T]$, bounded, and $\| B^{-1} A \| = K$.*

Proof For the self-adjoint compact linear operator B one can find an orthonormal eigenbasis $\{e_1, e_2, \ldots\}$ such that

$$B\left(\sum_{k=1}^{\infty} x_k e_k\right) = \sum_{k=1}^{\infty} \lambda_k x_k e_k \quad \text{for} \quad \sum_{k=1}^{\infty} x_k^2 < +\infty.$$

Then $\lim_{k\to\infty} \lambda_k = 0$ (by compactness), but for all k the inequality $\lambda_k \ne 0$ holds (by injectivity).

The inverse operator is a self-adjoint linear operator defined by the equation

$$B^{-1}\left(\sum_{k=1}^{\infty} x_k e_k\right) = \sum_{k=1}^{\infty} \frac{x_k}{\lambda_k} e_k \quad \text{for} \quad \sum_{k=1}^{\infty} \frac{x_k^2}{\lambda_k^2} < +\infty.$$

The domain of the operator B^{-1} is the subset

$$B(L_2[0, T]) = \left\{ \sum_{k=1}^{\infty} x_k e_k : \sum_{k=1}^{\infty} \frac{x_k^2}{\lambda_k^2} < \infty \right\}$$

of the Hilbert space $L_2[0, T]$.

Let us prove that the operator $B^{-1}A$ is defined on $L_2[0,T]$. Assume the opposite, i.e., $B^{-1}A$ is undefined at some point $f \in L_2[0,T]$. This means that $Af \notin B(L_2[0,T])$.

Decompose Af into a series by the eigenfunctions of the operator B:

$$Af = \sum_{k=1}^{\infty} x_k e_k. \tag{6.38}$$

Since $Af \notin B(L_2[0,T])$, we see that

$$\sum_{k=1}^{\infty} \frac{x_k^2}{\lambda_k^2} = +\infty,$$

and for

$$s_n = \sum_{k=1}^{n} \frac{x_k^2}{\lambda_k^2}$$

it holds that $\lim_{n\to\infty} s_n = +\infty$ and $s_n \geq 0$ for all $n \in \mathbb{N}$. Therefore, there exists $N \in \mathbb{N}$ such that

$$s_N > K^2 \|f\|^2. \tag{6.39}$$

Put

$$g = \sum_{k=1}^{N} \frac{x_k}{\lambda_k} e_k.$$

Then

$$\|g\|^2 = \sum_{k=1}^{N} \frac{x_k^2}{\lambda_k^2} = s_N, \qquad \|g\| = \sqrt{s_k}, \qquad g \in B(L_2[0,T]),$$

$$B^{-1}g = \sum_{k=1}^{N} \frac{x_k}{\lambda_k^2} e_k, \qquad \left\langle A^* B^{-1} g,\ f \right\rangle = \left\langle B^{-1}g, Af \right\rangle = \sum_{k=1}^{N} \frac{x_k}{\lambda_k^2} x_k = s_N.$$

By the Cauchy–Schwarz inequality,

$$\left|\left\langle A^* B^{-1} g,\ f \right\rangle\right| \leq \left\|A^* B^{-1} g\right\| \|f\| \leq \|A^* B^{-1}\| \|g\| \|f\| = K \sqrt{s_N}\, \|f\|.$$

Hence,

$$s_N \leq K \sqrt{s_N}\, \|f\|. \tag{6.40}$$

The inequalities (6.39) and (6.40) contradict each other. Thus, the operator $B^{-1}A$ is defined on the entire space $L_2[0,T]$.

Now let us prove boundedness of operator $B^{-1}A$ and the inequality $\left\|B^{-1}A\right\| \le K$. Suppose that this is not so. Then there exists an element f of the space $L_2[0,T]$ such that

$$\left\|B^{-1}Af\right\| > K\|f\|. \tag{6.41}$$

We use the same decomposition of the vector Af into the eigenvectors of B as above, see (6.38). Then

$$B^{-1}Af = \sum_{k=1}^{\infty} \frac{x_k}{\lambda_k} e_k, \qquad \left\|B^{-1}Af\right\|^2 = \sum_{k=1}^{\infty} \frac{x_k^2}{\lambda_k^2},$$
$$\lim_{n\to\infty} s_n = \left\|B^{-1}Af\right\|^2 > K^2\|f\|^2,$$

by (6.41). Therefore, there exists $N \in \mathbb{N}$ such that the inequality (6.39) holds. Arguing as above, we get a contradiction. Hence, $\|B^{-1}A\| \le K$.

It remains to prove the opposite inequality $\|B^{-1}A\| \ge K$. The operator A^*B^{-1} is defined on the set $B(L_2[0,T])$. For all $f \in B(L_2[0,T])$ from the domain of the operator A^*B^{-1}, we have

$$\|A^*B^{-1}f\|^2 = \langle B^{-1}AA^*B^{-1}f,\ f\rangle \le$$
$$\le \|B^{-1}AA^*B^{-1}f\|\,\|f\| \le \|B^{-1}A\|\,\|A^*B^{-1}f\|\,\|f\|,$$

whence

$$\|A^*B^{-1}f\| \le \|B^{-1}A\|\,\|f\|.$$

Therefore $K = \|A^*B^{-1}\| \le \|B^{-1}A\|$. □

6.6.4 The Proof of Boundedness and Compactness

Proposition 6.5 *Let* $\frac{1}{2} < H_1 < H_2 < 1$ *and* $H_1 \le \frac{3}{4}$. *Then* $\Gamma_{H_1}^{-1}\Gamma_{H_2}$ *is a compact linear operator defined on the entire space* $L_2[0,T]$.

Proof The operator Γ_{H_1} is an injective self-adjoint compact operator $L_2[0,T] \to L_2[0,T]$. The inverse operator $\Gamma_{H_1}^{-1}$ is densely defined on $L_2[0,T]$. By Proposition 6.4, the operators $I_{0+}^{2H_1-1}\Gamma_{H_1}^{-1}$ and $I_{T-}^{2H_1-1}\Gamma_{H_1}^{-1}$ are bounded. Therefore, by Lemma 6.1, the operators $\Gamma_{H_1}^{-1}I_{T-}^{2H_1-1}$ and $\Gamma_{H_1}^{-1}I_{0+}^{2H_1-1}$ are also bounded and defined on the entire space $L_2[0,T]$. By (6.34) and the semigroup property (Theorem 6.9),

$$\Gamma_{H_1}^{-1}\Gamma_{H_2} = H\Gamma(2H)\left(\Gamma_{H_1}^{-1}I_{0+}^{2H_2-1} + \Gamma_{H_1}^{-1}I_{T-}^{2H_2-1}\right) =$$
$$= H\Gamma(2H)\left(\Gamma_{H_1}^{-1}I_{0+}^{2H_1-1}I_{0+}^{2(H_2-H_1)} + \Gamma_{H_1}^{-1}I_{T-}^{2H_1-1}I_{T-}^{2(H_2-H_1)}\right).$$

Since $I_{0+}^{2(H_2-H_1)}$ and $I_{T-}^{2(H_2-H_1)}$ are compact operators, the operator $\Gamma_{H_1}^{-1}\Gamma_{H_2}$ is also compact. □

6.6.5 The Proof of Invertibility

Now prove that -1 is not an eigenvalue of the linear operator $\Gamma_{H_1}^{-1}\Gamma_{H_2}$. Indeed, if $\Gamma_{H_1}^{-1}\Gamma_{H_2} f = -f$ for some function $f \in L_2[0, T]$, then $\Gamma_{H_2} f + \Gamma_{H_1} f = 0$. Since Γ_{H_2} and Γ_{H_1} are positive definite self-adjoint (and injective) operators, $\Gamma_{H_2} + \Gamma_{H_1}$ is also a positive definite self-adjoint and injective operator. Hence $f = 0$ almost everywhere on $[0, T]$.

Because -1 is not an eigenvalue of the compact linear operator $\Gamma_{H_1}^{-1}\Gamma_{H_2}$, -1 is a regular point, i.e., $-1 \notin \sigma(\Gamma_{H_1}^{-1}\Gamma_{H_2})$, and the linear operator $\Gamma_{H_1}^{-1}\Gamma_{H_2} + I$ is invertible.

Acknowledgements The research of Yu. Mishura was funded (partially) by the Australian Government through the Australian Research Council (project number DP150102758). Yu. Mishura and K. Ralchenko acknowledge that the present research is carried through within the frame and support of the ToppForsk project nr. 274410 of the Research Council of Norway with title STORM: Stochastics for Time-Space Risk Models.

References

1. Belfadli, R., Es-Sebaiy, K., Ouknine, Y.: Parameter estimation for fractional Ornstein–Uhlenbeck processes: non-ergodic case. Front. Sci. Eng. **1**(1), 1–16 (2011)
2. Benassi, A., Cohen, S., Istas, J.: Identifying the multifractional function of a Gaussian process. Stat. Probab. Lett. **39**(4), 337–345 (1998)
3. Bercu, B., Coutin, L., Savy, N.: Sharp large deviations for the fractional Ornstein–Uhlenbeck process. Teor. Veroyatn. Primen. **55**(4), 732–771 (2010)
4. Berger, J., Wolpert, R.: Estimating the mean function of a Gaussian process and the Stein effect. J. Multivar. Anal. **13**(3), 401–424 (1983)
5. Bertin, K., Torres, S., Tudor, C.A.: Maximum-likelihood estimators and random walks in long memory models. Statistics **45**(4), 361–374 (2011)
6. Bojdecki, T., Gorostiza, L.G., Talarczyk, A.: Sub-fractional Brownian motion and its relation to occupation times. Stat. Probab. Lett. **69**(4), 405–419 (2004)
7. Cai, C., Chigansky, P., Kleptsyna, M.: Mixed Gaussian processes: a filtering approach. Ann. Probab. **44**(4), 3032–3075 (2016)
8. Cénac, P., Es-Sebaiy, K.: Almost sure central limit theorems for random ratios and applications to LSE for fractional Ornstein–Uhlenbeck processes. Probab. Math. Stat. **35**(2), 285–300 (2015)
9. Cheridito, P.: Mixed fractional Brownian motion. Bernoulli **7**(6), 913–934 (2001)
10. El Machkouri, M., Es-Sebaiy, K., Ouknine, Y.: Least squares estimator for non-ergodic Ornstein–Uhlenbeck processes driven by Gaussian processes. J. Korean Stat. Soc. **45**(3), 329–341 (2016)
11. Es-Sebaiy, K.: Berry–Esséen bounds for the least squares estimator for discretely observed fractional Ornstein–Uhlenbeck processes. Stat. Probab. Lett. **83**(10), 2372–2385 (2013)
12. Es-Sebaiy, K., Ndiaye, D.: On drift estimation for non-ergodic fractional Ornstein–Uhlenbeck process with discrete observations. Afr. Stat. **9**(1), 615–625 (2014)

13. Es-Sebaiy, K., Ouassou, I., Ouknine, Y.: Estimation of the drift of fractional Brownian motion. Stat. Probab. Lett. **79**(14), 1647–1653 (2009)
14. Houdré, C., Villa, J.: An example of infinite dimensional quasi-helix. In: Stochastic Models: Seventh Symposium on Probability and Stochastic Processes, 23–28 June 2002, Mexico City, Mexico. Selected papers, pp. 195–201. American Mathematical Society (AMS), Providence, RI (2003)
15. Hu, Y., Nualart, D.: Parameter estimation for fractional Ornstein–Uhlenbeck processes. Stat. Probab. Lett. **80**(11–12), 1030–1038 (2010)
16. Hu, Y., Song, J.: Parameter estimation for fractional Ornstein–Uhlenbeck processes with discrete observations. In: Malliavin calculus and stochastic analysis. A Festschrift in honor of David Nualart, pp. 427–442. Springer, New York (2013)
17. Hu, Y., Nualart, D., Xiao, W., Zhang, W.: Exact maximum likelihood estimator for drift fractional Brownian motion at discrete observation. Acta Math. Sci. Ser. B Engl. Ed. **31**(5), 1851–1859 (2011)
18. Hu, Y., Nualart, D., Zhou, H.: Parameter estimation for fractional Ornstein–Uhlenbeck processes of general Hurst parameter. arXiv preprint, arXiv:1703.09372 (2017)
19. Ibragimov, I.A., Rozanov, Y.A.: Gaussian Random Processes. Applications of Mathematics, vol. 9. Springer, New York (1978)
20. Kleptsyna, M.L., Le Breton, A.: Statistical analysis of the fractional Ornstein–Uhlenbeck type process. Stat. Inference Stoch. Process. **5**, 229–248 (2002)
21. Kozachenko, Y., Melnikov, A., Mishura, Y.: On drift parameter estimation in models with fractional Brownian motion. Statistics **49**(1), 35–62 (2015)
22. Kubilius, K., Mishura, Y.: The rate of convergence of Hurst index estimate for the stochastic differential equation. Stoch. Process. Appl. **122**(11), 3718–3739 (2012)
23. Kubilius, K., Mishura, Y., Ralchenko, K., Seleznjev, O.: Consistency of the drift parameter estimator for the discretized fractional Ornstein–Uhlenbeck process with Hurst index $H \in (0, \frac{1}{2})$. Electron. J. Stat. **9**(2), 1799–1825 (2015)
24. Kubilius, K.e., Mishura, Y., Ralchenko, K.: Parameter Estimation in Fractional Diffusion Models. Bocconi & Springer Series, vol. 8. Bocconi University Press, Milan; Springer, Cham (2017)
25. Kukush, A., Mishura, Y., Valkeila, E.: Statistical inference with fractional Brownian motion. Stat. Inference Stoch. Process. **8**(1), 71–93 (2005)
26. Le Breton, A.: Filtering and parameter estimation in a simple linear system driven by a fractional Brownian motion. Stat. Probab. Lett. **38**(3), 263–274 (1998)
27. Mishura, Y.: Stochastic Calculus for Fractional Brownian Motion and Related Processes, vol. 1929. Springer Science & Business Media (2008)
28. Mishura, Y.: Maximum likelihood drift estimation for the mixing of two fractional Brownian motions. In: Stochastic and Infinite Dimensional Analysis, pp. 263–280. Springer, Berlin (2016)
29. Mishura, Y., Ralchenko, K.: On drift parameter estimation in models with fractional Brownian motion by discrete observations. Austrian J. Stat. **43**(3), 218–228 (2014)
30. Mishura, Y., Voronov, I.: Construction of maximum likelihood estimator in the mixed fractional-fractional Brownian motion model with double long-range dependence. Mod. Stoch. Theory Appl. **2**(2), 147–164 (2015)
31. Mishura, Y., Ralchenko, K.: Drift parameter estimation in the models involving fractional Brownian motion. In: Panov, V. (ed.) Modern Problems of Stochastic Analysis and Statistics: Selected Contributions in Honor of Valentin Konakov, pp. 237–268. Springer International Publishing, Cham (2017)
32. Mishura, Y., Ralchenko, K., Seleznev, O., Shevchenko, G.: Asymptotic properties of drift parameter estimator based on discrete observations of stochastic differential equation driven by fractional Brownian motion. In: Modern stochastics and Applications. Springer Optimization and Its Applications, vol. 90, pp. 303–318. Springer, Cham (2014)
33. Mishura, Y., Ralchenko, K., Shklyar, S.: Maximum likelihood drift estimation for Gaussian process with stationary increments. Austrian J. Stat. **46**(3–4), 67–78 (2017)
34. Mishura, Y., Ralchenko, K., Shklyar, S.: Maximum likelihood drift estimation for Gaussian process with stationary increments. Nonlinear Anal. Model. Control **23**(1), 120–140 (2018)

35. Norros, I., Valkeila, E., Virtamo, J.: An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. Bernoulli **5**(4), 571–587 (1999)
36. Peltier, R.F., Lévy Véhel, J.: Multifractional Brownian motion: definition and preliminary results. INRIA Research Report, vol. 2645 (1995)
37. Polyanin, A., Manzhirov, A.: Handbook of Integral Equations, 2nd edn. Chapman & Hall/CRC, Boca Raton (2008)
38. Prakasa Rao, B.L.S.: Statistical Inference for Fractional Diffusion Processes. Wiley, Chichester (2010)
39. Privault, N., Réveillac, A.: Stein estimation for the drift of Gaussian processes using the Malliavin calculus. Ann. Stat. **36**(5), 2531–2550 (2008)
40. Ralchenko, K.V., Shevchenko, G.M.: Paths properties of multifractal Brownian motion. Theory Probab. Math. Stat. **80**, 119–130 (2010)
41. Samko, S., Kilbas, A., Marichev, O.: Fractional Integrals and Derivatives. Taylor & Francis (1993)
42. Samuelson, P.A.: Rational theory of warrant pricing. Ind. Manag. Rev. **6**(2), 13–32 (1965)
43. Shen, G., Yan, L.: Estimators for the drift of subfractional Brownian motion. Commun. Stat. Theory Methods **43**(8), 1601–1612 (2014)
44. Tanaka, K.: Distributions of the maximum likelihood and minimum contrast estimators associated with the fractional Ornstein–Uhlenbeck process. Stat. Inference Stoch. Process. **16**, 173–192 (2013)
45. Tanaka, K.: Maximum likelihood estimation for the non-ergodic fractional Ornstein–Uhlenbeck process. Stat. Inference Stoch. Process. **18**(3), 315–332 (2015)
46. Tudor, C.: Some properties of the sub-fractional Brownian motion. Stochastics **79**(5), 431–448 (2007)
47. Tudor, C.A., Viens, F.G.: Statistical aspects of the fractional stochastic calculus. Ann. Stat. **35**(3), 1183–1212 (2007)
48. Xiao, W., Zhang, W., Xu, W.: Parameter estimation for fractional Ornstein–Uhlenbeck processes at discrete observation. Appl. Math. Model. **35**, 4196–4207 (2011)
49. Xiao, W.L., Zhang, W.G., Zhang, X.L.: Maximum-likelihood estimators in the mixed fractional Brownian motion. Statistics **45**, 73–85 (2011)
50. Zabreyko, P.P., Koshelev, A.I., Krasnosel'skii, M.A., Mikhlin, S.G., Rakovshchik, L.S., Stet'senko, V.Y.: Integral Equations: A Reference Text. Noordhoff, Leyden (1975)

Chapter 7
Application of Limit Theorems for Superposition of Random Functions to Sequential Estimation

Gulnoza Rakhimova

Abstract The paper presents sequential fixed-width confidence interval estimators for functionals of an unknown distribution function. Conditions of asymptotic consistency for fixed-width confidence interval estimators and asymptotic efficiency of stopping times are given.

Keywords Sequential estimation · Stopping time · Fixed-width confidence interval · Asymptotic consistency · Asymptotic efficiency

7.1 Introduction

Beginning from the originating works [3, 4] many authors used the classical Anskombe theorem presented in [2] for proofs of asymptotic consistency for confidence intervals with fixed width for different concrete models. The above theorem assume holding of so-called condition of uniform continuity in probability. The survey of related works is given in book [5].

An alternative approach to the proof of asymptotic consistency for confidence intervals with fixed width, based on the weak invariance principle has been developed in [11]. A number of concrete models, where the proof of asymptotic consistency for confidence intervals with fixed width is based on the use of general functional limit theorems for randomly stopped stochastic processes, were presented in [12, 13], and in papers [1, 6–10].

In the present paper, we present this method and the corresponding results in the more general form.

G. Rakhimova (✉)
Tashkent Auto-Road Institute, Tashkent, Uzbekistan
e-mail: rakhimova.gulnoza@mail.ru

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_7

7.2 Asymptotic Consistency and Efficiency of Confidence Intervals with Fixed Width

Let $\xi_1, \ldots, \xi_n$ be real-valued i.i.d. random variables with unknown distribution function $F(x)$.

Let, also, $\Theta(F)$ be some real-valued functional of this distribution function. In order to estimate this functional, we consider a statistical estimators $\Theta_n(F) = \Theta_n(\xi_1, \ldots, \xi_n)$, which have finite expectations and, therefore, can be represented in the following form,

$$\Theta_n(F) = \Theta(F) + (\Theta_n(F) - \mathsf{E}\Theta_n(F)) + (\mathsf{E}\Theta_n(F)) - \Theta(F)). \tag{7.1}$$

We assume that the random variables $\Theta_n(F) - \mathsf{E}\Theta_n(F)$ admit the following representation, for every $n \geq 1$,

$$\Theta_n(F) - \mathsf{E}\Theta_n(F) = \frac{1}{n}\sum_{k=1}^{n} b_n(F) Y(F, \xi_k), \tag{7.2}$$

where $Y(F, \xi_k), k = 1, \ldots, n$ are i.i.d. random variables ($Y(F, \cdot)$ is a real-valued Borel function defined on $\mathbb{R}_1$) and $b_n(F)$ are real-valued constants, which satisfy the following condition:

A: **(a)** $\mathsf{E}Y(F, \xi_1) = 0$, **(b)** $b_n(F) \to 1$ as $n \to \infty$, **(c)** $n^{-\alpha}\sum_{k=1}^{n} Y(F, \xi_k) \xrightarrow{d} Y(F)$ as $n \to \infty$, where **(d)** $\frac{1}{2} \leq \alpha < 1$, **(e)** $c(F)$ is a positive constant, **(f)** $Y(F)$ is a symmetric stable random variable with characteristic function $\mathsf{E}e^{\mathbf{i}sY(F)} = e^{-|s|^{\frac{1}{\alpha}}c(F)}$, $s \in \mathbb{R}_1$.

Symbol $\xrightarrow{d}$ denotes convergence in distribution of random variables.

It is appropriate to mention an important case of condition **A**, where $\alpha = \frac{1}{2}$ and $Y(F)$ is the standard normal random variable with mean 0 and variance $c(F)$.

Let us denote, for $n \geq 1$,

$$B_n(F) = \mathsf{E}\Theta_n(F) \text{ and } Z_n(F) = B_n(F) - \Theta(F). \tag{7.3}$$

The non-random quantity $Z_n(F)$ is the expected bias of estimator $\theta_n(F)$.

We also assume that the following condition holds:

B: $n^{1-\alpha} Z_n(F) \to 0$ as $n \to \infty$.

An important example of the model described above is where the functional $\Theta(F) = \mathsf{E}f(\xi_1) = \int_{-\infty}^{\infty} f(x)F(dx)$, where $f(x)$ is a real-valued function absolutely integrable with respect to probability measure $F(dx)$.

In this case, $\Theta_n(F) = \frac{1}{n}\sum_{k=1}^{n} b_n(F) f(\xi_k)$ is a natural choice for the estimator $\Theta_n(F)$. Respectively, random variables $Y(F, \xi_k) = f(\xi_k) - \mathsf{E}f(\xi_k) = f(\xi_k) -$

$\Theta(F), k \geq 1$, and, thus, $\mathsf{E}\Theta_n(F) = b_n(F)\Theta(F)$, and $Z_n(F) = (b_n(F) - 1)\Theta(F)$. Constant $b_n(F) - 1$ is, in this case, a multiplicative bias factor.

Let $\Phi_\alpha(x)$ be the distribution function of the stable random variable with the characteristic function $e^{-|s|^{\frac{1}{\alpha}}}$. It is a strictly monotonic continuous function. Let also $0 < \gamma < 1$ and $a_\gamma = \Phi_\alpha^{-1}(\frac{1+\gamma}{2})$.

Let us define, for every $\varepsilon > 0$ and $n \geq 1$, the random interval of the length 2ε,

$$\mathrm{I}(n) = [\Theta_n(F) - \varepsilon, \Theta_n(F) + \varepsilon], \tag{7.4}$$

and

$$n_{\gamma,\varepsilon} = \min\left(n \geq 1 : n > \left(\frac{a_\gamma c(F)^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right). \tag{7.5}$$

Relation (7.5) permits obviously to write down for $n_{\gamma,\varepsilon}$ the alternative formula,

$$n_{\gamma,\varepsilon} = \left[\left(\frac{a_\gamma c(F)^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right] + 1 \geq \left(\frac{a_\gamma c(F)^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}, \tag{7.6}$$

where $[x]$ denote the integer part of real number x.

Conditions **A**, **B** and Slutsky theorem (see, for example, [13]) imply that the following relation holds for every $0 < \gamma < 1$,

$$\begin{aligned}
\mathsf{P}\{\Theta(F) \in \mathrm{I}(n_{\gamma,\varepsilon})\} &= \mathsf{P}\{|\Theta_{n_{\gamma,\varepsilon}}(F) - \Theta(F)| \leq \varepsilon\} \\
&\geq \mathsf{P}\{|\Theta_{n_{\gamma,\varepsilon}}(F) - B_{n_{\gamma,\varepsilon}}(F)| + |Z_{n_{\gamma,\varepsilon}}| \leq \varepsilon\} \\
&= \mathsf{P}\left\{|c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha}(\Theta_{n_{\gamma,\varepsilon}}(F) - B_{n_{\gamma,\varepsilon}}(F))|\right. \\
&\qquad \left. + |c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha} Z_{n_{\gamma,\varepsilon}}| \leq \frac{c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha}\varepsilon}{a_\gamma} a_\gamma\right\} \\
&\geq \mathsf{P}\{|b_{n_{\gamma,\varepsilon}}(F) c(F)^{-\alpha} n_{\gamma,\varepsilon}^{-\alpha} \sum_{k=1}^{n} Y(F, \xi_k)| \\
&\qquad + |c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha} Z_{n_{\gamma,\varepsilon}}| \leq a_\gamma\} \\
&\to 2\Phi_\alpha(a_\gamma) - 1 = \gamma \text{ as } \varepsilon \to 0.
\end{aligned} \tag{7.7}$$

Relation (7.7) means that $\mathrm{I}(n_{\gamma,\varepsilon})$ is an asymptotic confidence interval of level γ for the unknown functional $\Theta(F)$, for every $0 < \gamma < 1$.

The sequential procedure described above has the serious shortage. It depends on the unknown functional $c(F)$.

In order to improve the above sequential procedure, we should replace $c(F)$ by some consistent estimator of this functional.

Let $V_n = V_n(\xi_1, \ldots, \xi_n)$, $n \geq 1$ be a.s. positive consistent estimators of functional $c(F)$, i.e. this estimators satisfy the following condition:

C: **(a)** $\mathsf{P}\{V_n > 0\} = 1, n \geq 1$, **(b)** $V_n \xrightarrow{a.s.} c(F)$ as $n \to \infty$.

Symbol $\xrightarrow{a.s.}$ denotes almost sure convergence of random variables.

Let us now introduce the random stopping moments,

$$N_{\gamma,\varepsilon} = \min\left(n \geq 1 : n > \left(\frac{a_\gamma V_n^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right). \tag{7.8}$$

and consider the system of confidence intervals $\mathrm{I}(N_{\gamma,\varepsilon})$, $0 < \gamma < 1, \varepsilon > 0$.

It should be noted that a relation analogous to (7.6) does not take place for random stopping times $N_{\gamma,\varepsilon}$ since variability of values $V_n, n \geq 1$. Such, relation would, however, take place for the case if all random variables $V_n, n \geq 1$ would be replaced by some random variable V in the defining relation (7.8).

Lemma 7.1 *Let condition* **C** *holds. Then:* **(i)** $\mathsf{P}\{N_{\gamma,\varepsilon}) < \infty$, *for every* $0 < \gamma < 1, \varepsilon > 0$; **(ii)** $N_{\gamma,\varepsilon} \xrightarrow{a.s.} \infty$ *as* $\varepsilon \to 0$, *for every* $0 < \gamma < 1$.

Proof Relation (7.8) implies that the following relation holds, for every $0 < \gamma < 1, \varepsilon > 0$ and $n \geq 1$,

$$\begin{aligned}\mathsf{P}\{N_{\gamma,\varepsilon} > n\} &= \mathsf{P}\left\{k \leq \left(\frac{a_\gamma V_k^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}, k = 1, \ldots, n\right\} \\ &\leq \mathsf{P}\left\{n \leq \left(\frac{a_\gamma V_n^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right\}. \end{aligned} \tag{7.9}$$

Let $\langle \Omega, \mathcal{F}, \mathsf{P}\rangle$ be the probability space on which the sequence of random variables $\xi_n, n \geq 1$ is defined. A be the set of elementary events ω, for which $V_n(\omega) \to c(F)$ as $n \to \infty$ and B the set of elementary events ω, for which $V_n(\omega) > 0, n \geq 1$. The assumption of a.s. positivity of the random variables $V_n, n \geq 1$ and condition **C** implies that $\mathsf{P}(A \cap B) = 1$.

Let us define the random variable $U_+ = \max_{n\geq 1} V_n$. Obviously, $U_+(\omega) < \infty$, for $\omega \in A$, and, thus, $\mathsf{P}\{U_+ < \infty\} = 1$. Therefore,

$$\begin{aligned}\lim_{n\to\infty} \mathsf{P}\{N_{\gamma,\varepsilon} > n\} &\leq \lim_{n\to\infty} \mathsf{P}\left\{n < \left(\frac{a_\gamma V_n^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right\} \\ &\leq \lim_{n\to\infty} \mathsf{P}\left\{n < \left(\frac{a_\gamma U_+^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right\} = 0. \end{aligned} \tag{7.10}$$

Relation (7.10) proves proposition **(i)** of Lemma 7.1.

Let us also define the random variable $U_- = \min_{n\geq 1} V_n$. Obviously, $0 < U_-(\omega) < \infty$, for $\omega \in A \cap B$, and, thus, $\mathsf{P}\{0 < U_- < \infty\} = 1$. Therefore,

$$N_{\gamma,\varepsilon} \geq \min\left(n \geq 1 : n > \left(\frac{a_\gamma U_-^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right) \xrightarrow{a.s.} \infty \text{ as } \varepsilon \to 0. \tag{7.11}$$

Relation (7.11) proves proposition **(ii)** of Lemma 7.1. □

The system of confidence intervals $\mathrm{I}(N_{\gamma,\varepsilon}), 0 < \gamma < 1, \varepsilon > 0$ for the functional $\Theta(F)$ is asymptotically consistent, if the following relation holds, for every $0 < \gamma < 1$,

$$\underline{\lim}_{\varepsilon\to 0} \mathsf{P}\{\Theta(F) \in \mathrm{I}(N_{\gamma,\varepsilon})\} \geq \gamma. \tag{7.12}$$

The system of asymptotically consistent confidence intervals $\mathrm{I}(N_{\gamma,\varepsilon}), 0 < \gamma < 1, \varepsilon > 0$ is asymptotically a.s. efficient if, together with relation (7.12), the following relation holds, for every $0 < \gamma < 1$,

$$\frac{N_{\gamma,\varepsilon}}{n_{\gamma,\varepsilon}} \xrightarrow{a.s.} 1 \text{ as } \varepsilon \to 0. \tag{7.13}$$

Theorem 7.1 *Let conditions* **A**–**C** *hold. Then,* $\mathrm{I}(N_{\gamma,\varepsilon}), 0 < \gamma < 1, \varepsilon > 0$ *is asymptotically consistent and a.s. efficient system of confidence intervals with fixed width, for the the functional* $\Theta(F)$.

Proof Let A and B be the random events defined in the proof of Lemma 7.1. Let us choose an arbitrary elementary event $\omega \in A \cap B$ and $0 < \delta < c(F)$.

Let us also define,

$$m_\delta(\omega) = \max(n \geq 1 : |V_n(\omega) - c(F)| \geq \delta), \tag{7.14}$$

and

$$\varepsilon_\delta(\omega) = \sup\left(\varepsilon > 0 : k \leq \left(\frac{a_\gamma V_k^\alpha(\omega)}{\varepsilon}\right)^{\frac{1}{1-\alpha}}, 1 \leq k \leq m_\delta(\omega)\right). \tag{7.15}$$

Relations (7.14) and (7.15) imply that, for $0 < \varepsilon < \varepsilon_\delta(\omega)$,

$$\begin{aligned} N_{\gamma,\varepsilon}(\omega) &= \min\left(n \geq 1 : n > \left(\frac{a_\gamma V_n^\alpha(\omega)}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right) \\ &= \min\left(n \geq m_\delta(\omega) : n > \left(\frac{a_\gamma V_n^\alpha(\omega)}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right) \\ &\geq \min\left(n \geq m_\delta(\omega) : n > \left(\frac{a_\gamma (c(F)-\delta)^\alpha}{\varepsilon}\right)^{\frac{1}{1-\alpha}}\right) \end{aligned}$$

$$= \left[\left(\frac{a_\gamma (c(F) - \delta)^\alpha}{\varepsilon} \right)^{\frac{1}{1-\alpha}} \right] + 1. \tag{7.16}$$

Analogously, relations (7.14) and (7.15) imply that, for any $0 < \varepsilon < \varepsilon_\delta(\omega)$,

$$N_{\gamma,\varepsilon}(\omega) \le \left[\left(\frac{a_\gamma (c(F) + \delta)^\alpha}{\varepsilon} \right)^{\frac{1}{1-\alpha}} \right] + 1. \tag{7.17}$$

Relations (7.6), (7.16) and (7.17) imply that the following two-sided inequalities hold, for $0 < \varepsilon < \varepsilon_\delta(\omega)$,

$$\frac{\left[\left(\frac{a_\gamma (c(F)-\delta)^\alpha}{\varepsilon} \right)^{\frac{1}{1-\alpha}} \right] + 1}{\left[\left(\frac{a_\gamma c(F)^\alpha}{\varepsilon} \right)^{\frac{1}{1-\alpha}} \right] + 1} \le \frac{N_{\gamma,\varepsilon}(\omega)}{n_{\gamma,\varepsilon}} \le \frac{\left[\left(\frac{a_\gamma (c(F)+\delta)^\alpha}{\varepsilon} \right)^{\frac{1}{1-\alpha}} \right] + 1}{\left[\left(\frac{a_\gamma c(F)^\alpha}{\varepsilon^2} \right)^{\frac{1}{1-\alpha}} \right] + 1}. \tag{7.18}$$

Relation (7.18) implies in an obvious way that,

$$\left(\frac{c(F) - \delta}{c(F)} \right)^{\frac{\alpha}{1-\alpha}} \le \underline{\lim}_{\varepsilon \to 0} \frac{N_{\gamma,\varepsilon}(\omega)}{n_{\gamma,\varepsilon}} \le \overline{\lim}_{\varepsilon \to 0} \frac{N_{\gamma,\varepsilon}(\omega)}{n_{\gamma,\varepsilon}} \le \left(\frac{c(F) + \delta}{c(F)} \right)^{\frac{\alpha}{1-\alpha}}. \tag{7.19}$$

Since, an arbitrary choice of $0 < \delta < c(F)$, relation (7.19) implies that,

$$\frac{N_{\gamma,\varepsilon}(\omega)}{n_{\gamma,\varepsilon}} \to 1 \text{ as } \varepsilon \to 0. \tag{7.20}$$

Since, an arbitrary choice of $\omega \in A \cap B$, relation (7.20) implies that relation (7.13) holds.

Let us define, for every $1 < \gamma < 1$, $\varepsilon > 0$, the sum-process,

$$Y_{\gamma,\varepsilon}(t) = c(F)^{-\alpha} n_{\gamma,\varepsilon}^{-\alpha} \sum_{k \le t n_{\gamma,\varepsilon}} Y(F, \xi_k), \ t \ge 0. \tag{7.21}$$

As was shown in [14], condition **A** implies that,

$$Y_{\gamma,\varepsilon}(t), t \ge 0 \xrightarrow{\mathsf{J}} W_\alpha(t), t \ge 0 \text{ as } \varepsilon \to 0, \tag{7.22}$$

where $W_\alpha(t), t \ge 0$ is a càdlàg Lévy process with the characteristic functions $\mathsf{E} e^{\mathbf{i} s W_\alpha(t)} = e^{-|s|^{\frac{1}{\alpha}} t}$, $s \in \mathbb{R}_1$, $t \ge 0$.

Here, symbol $\xrightarrow{\mathsf{J}}$ denotes convergence in Skorokhod J-topology for càdlàg stochastic processes.

Relations (7.13) and (7.22) imply, by Slutsky theorem, that for every $0 < \gamma < 1$,

$$\left(\frac{N_{\gamma,\varepsilon}}{n_{\gamma,\varepsilon}}, Y_{\gamma,\varepsilon}(t)\right), t \geq 0 \Longrightarrow (1, W_\alpha(t)), t \geq 0 \text{ as } \varepsilon \to 0, \tag{7.23}$$

Here, symbol $\Longrightarrow$ is used to denote weak convergence of finite-dimensional distributions for stochastic processes.

As it follows from results given in [13, Theorem 2.2.1], relations (7.22) and (7.23) imply that the following relation holds,

$$Y_{\gamma,\varepsilon}\left(\frac{N_{\gamma,\varepsilon}}{n_{\gamma,\varepsilon}}\right) = c(F)^{-\alpha} n_{\gamma,\varepsilon}^{-\alpha} \sum_{k \leq N_{\gamma,\varepsilon}} Y_n(F, \xi_k) \xrightarrow{d} W_\alpha(1) \text{ as } \varepsilon \to 0. \tag{7.24}$$

Condition **A (a)** and proposition **(ii)** of Lemma 7.1 imply the following relation,

$$b_{N_{\gamma,\varepsilon}} \xrightarrow{a.s} 1 \text{ as } \varepsilon \to 0. \tag{7.25}$$

Relations (7.13), (7.24) and (7.25) imply, by Slutsky theorem, that for every $0 < \gamma < 1$,

$$\begin{aligned} c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha} (\Theta_{N_{\gamma,\varepsilon}} - B_{N_{\gamma,\varepsilon}}) \\ &= \left(\frac{N_{\gamma,\varepsilon}}{n_{\gamma,\varepsilon}}\right)^{-1} b_{N_{\gamma,\varepsilon}} c(F)^{-\alpha} n_{\gamma,\varepsilon}^{-\alpha} \sum_{k \leq N_{\gamma,\varepsilon}} Y_n(F, \xi_k) \\ &\xrightarrow{d} 1 \cdot 1 \cdot W_\alpha(1) = W_\alpha(1) \text{ as } \varepsilon \to 0. \end{aligned} \tag{7.26}$$

Also, relation (7.13), condition **B**, and Lemma 7.1 imply in obvious way that,

$$n_{\gamma,\varepsilon}^{1-\alpha} Z_{N_{\gamma,\varepsilon}} = \left(\frac{N_{\gamma,\varepsilon}}{n_{\gamma,\varepsilon}}\right)^{\alpha-1} \cdot N_{\gamma,\varepsilon}^{1-\alpha} Z_{N_{\gamma,\varepsilon}} \xrightarrow{a.s} 1 \cdot 0 = 0 \text{ as } \varepsilon \to 0. \tag{7.27}$$

Using relations (7.26), (7.27), and Slutsky theorem, we get the following relation, for every $0 < \gamma < 1$,

$$\begin{aligned} \mathsf{P}\{\Theta(F) \in \mathrm{I}(N_{\gamma,\varepsilon})\} &= \mathsf{P}\{|\Theta_{N_{\gamma,\varepsilon}}(F) - \Theta(F)| \leq \varepsilon\} \\ &\geq \mathsf{P}\{|\Theta_{N_{\gamma,\varepsilon}}(F) - B_{N_{\gamma,\varepsilon}}(F)| + |Z_{N_{\gamma,\varepsilon}}| \leq \varepsilon\} \\ &= \mathsf{P}\Big\{|c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha} (\Theta_{N_{\gamma,\varepsilon}}(F) - B_{N_{\gamma,\varepsilon}}(F))| \\ &\qquad + |c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha} Z_{N_{\gamma,\varepsilon}}| \leq \frac{c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha} \varepsilon}{a_\gamma} a_\gamma\Big\} \\ &\geq \mathsf{P}\{|c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha} (\Theta_{N_{\gamma,\varepsilon}}(F) - B_{N_{\gamma,\varepsilon}}(F))| \\ &\qquad + |c(F)^{-\alpha} n_{\gamma,\varepsilon}^{1-\alpha} Z_{N_{\gamma,\varepsilon}}| \leq a_\gamma\} \\ &\to 2\Phi_\alpha(a_\gamma) - 1 = \gamma \text{ as } \varepsilon \to 0. \end{aligned} \tag{7.28}$$

Thus, relation (7.12) holds. □

References

1. Abdushukurov, A., Tursunov, G., Rakhimova, G.: Sequential Estimation by Intervals of Fixed Width. Lambert Academic Publishing, Germany (2017)
2. Anscombe, F.J.: Large sample theory of sequential estimation. Proc. Camb. Philos. Soc. **48**, 600–607 (1952)
3. Chow, Y.S., Robbins, H.: On the asymptotic theory of fixed-width sequential confidence intervals for the mean. Ann. Math. Stat. **36**, 457–462 (1965)
4. Chow, Y.S., Robbins, H.: Great Expectations: The Theory of Optimal Stopping. Houghton Mafflin, Boston (1971)
5. Ghosh, M., Mukhopadhyay, N., Sen, P.K.: Sequential Estimation. Wiley, New York (1996)
6. Rakhimova, G.G., Tursunov, G.T.: Sequential estimation of the transfer factor in a differential equation by intervals of fixed width. Materials of Republic Scientific and Practical Conference "Statistics and Its Applications". Tashkent, pp. 74–77 (2013)
7. Rakhimova, G.G., Tursunov, G.T.: Sequential non-parametric estimation by intervals of fixed width of a regression function. Materials of International Scientific Conference "Theory of Probability, Stochastic Processes, Mathematical Statistics and Applications". Minsk, pp. 259–261 (2015)
8. Rakhimova, G.G., Tursunov, G.T.: Sequential estimation of by intervals of fixed width of mean value. Statistical Methods of Estimation and Hypotheses Testing. Intercollegiate Collection of Scientific Papers. Perm, vol. 26, pp. 58–67 (2015)
9. Rakhimova, G.G., Tursunov, G.T.: Sequential estimation of by intervals of fixed width for nonparametric statistics. Problems of Modern Topology and Applications. Tashkent, pp. 248–249 (2017)
10. Rakhimova, G.G., Siguldaev, K.S., Tursunov, G.T.: Rate of convergence in sequential estimation by intervals of fixed width of asymptotic variance for rank estimators of shift parameter. In: Proceedings of XIII Conference on Financial and Actuarial Mathematics and Multivariate Statistics. Krasnoyarsk, pp. 202–203 (2015)
11. Sen, P.K.: Sequential Nonparametrics: Invariance Principles and Statistical Inference. Wiley, New York (1981)
12. Silvestrov, D.S.: Limit Theorems for Composite Random Functions. Vysshaya Shkola and Izdatel'stvo Kievskogo Universiteta, Kiev (1974)
13. Silvestrov, D.S.: Limit Theorems for Randomly Stopped Stochastic Processes. Probability and Its Applications. Springer, London (2004)
14. Skorokhod, A.V.: Limit theorems for processes with independent increments. Teor. Veroyatn. Primen. **2**, 145–177 (1957) (English translation in Theory Probab. Appl. **2**, 138–171)

Chapter 8
On Simulation of a Fractional Ornstein–Uhlenbeck Process of the Second Kind by the Circulant Embedding Method

José Igor Morlanes and Andriy Andreev

Abstract We demonstrate how to utilize the Circulant Embedding Method (*CEM*) for simulation of fractional Ornstein–Uhlenbeck process of the second kind (fOU_2). The algorithm contains two major steps. First, the relevant covariance matrix is embedded into a circulant one. Second, a sample from the fOU_2 is obtained by means of fast Fourier transform applied on the circulant extended matrix. The main goal of this paper is to explain both steps in detail. As a result, we obtain an accurate and an efficient algorithm for generating fOU_2 random vectors. We also indicate that the above described procedure can be extended to applications with non-Gaussian marginals.

Keywords Fractional Brownian motion · Fractional Ornstein–Uhlenbeck process · Circulant embedding method · Simulation

8.1 Introduction

Fractional Ornstein–Uhlenbeck processes of the second kind (fOU_2) comprises a family of Gaussian processes constructed via Lamperti transform of fractional Brownian motion. It was introduced by Kaarakka and Salminen [13] and further studied by Azmoodeh and Morlanes [4] as well as by Azmoodeh and Viitari [5]. The main appeal of the fOU_2 is that it is a short range dependent process for all values of H. Contrasted to the fact that *fBM* is long range dependent for $H \in (\frac{1}{2}, 1)$, the fOU_2 becomes an interesting object for applications in e.g. physics and finance.

Among existing algorithms for simulation of stochastic processes with a given covariance structure, Hosking's method [12] and Cholesky decomposition [2] are

J. I. Morlanes · A. Andreev (✉)
Department of Statistics, Stockholm University, Stockholm, Sweden
e-mail: andriy.andreev@stat.su.se

J. I. Morlanes
e-mail: jose.morlanes@stat.su.se

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_8

the standard choices. These methods can be adapted to accurately simulate one-dimensional fOU_2, but all of them have a generic drawback of high computational cost, when realised on a fine grid since the covariance matrix becomes too large and causes computational problems. However, if the covariance matrix is circulant, there is a more efficient and accurate way to progress with simulations. Circulant embedding methods were introduced by Davies and Harte [8] and generalised by Wood and Chan [6, 7, 18], Dietrich and Newsam [9], and Gneiting et al. [10].

The purpose of this paper is to describe how to adapt the circulant embedding method for simulation of fOU_2. Although the fOU_2 covariance matrix is not circulant, we can embed it into a circulant matrix. We demonstrate how to do so for one-dimensional fOU_2 random vectors on a discrete grid. Similar techniques can be extended to the multidimensional case. Further, one can extend these methods to non-Gaussian marginals, see Grigoriu [11]. We show that the algorithm works efficiently and then demonstrate sub-steps of the simulation algorithm explicitly on simple examples, employing fast Fourier transform to speed up computations.

This paper is organised as follows: Sect. 8.2 gives a brief summary of basic properties of fractional Brownian motion. Section 8.3 introduces fOU_2 and describes its covariance function. Circulant embedding method and practicalities of the generation of fOU_2 vector are given and then followed by an example in Sect. 8.4. We conclude and provide some discussion of the findings in Sect. 8.5.

8.2 Fractional Brownian Motion

The fractional Brownian motion serves as a good benchmark and a starting point in the construction of an Ornstein–Uhlenbeck process of the second kind. The term "fractional" comes due to Mandelbrot and van Ness [14] and they were the first to establish the integral representation of the process. Nualart [15] describes most of the properties we are interested in. The following section reviews the most relevant of them, that might later be contrasted with properties of fOU_2.

8.2.1 Definition and Some Basic Properties of fBM

A fractional Brownian motion (fBM) is defined as a continuous Gaussian process $B_t^H, t \geq 0$ with $B_0^H = 0$ with covariance function

$$Cov(B_t^H, B_s^H) = \frac{1}{2}\sigma^2(t^{2H} + s^{2H} - (t-s)^{2H}) \tag{8.1}$$

for all $0 \leq s \leq t$ and Hurst parameter $(0 < H < 1)$. In particular, $Var(B_t^H) = \sigma^2 t^{2H}$, and if $H = \frac{1}{2}$, the variance becomes $\sigma^2 t$ which corresponds to Brownian motion. Self-similarity is another key property that immediately follows from (8.1). Increments $(B_t^H - B_s^H)$ are Gaussian stationary with mean zero and variance $Var(B_t^H - B_s^H) = \sigma^2|t-s|^{2H}$, the latter being the starting point for simulation of

fBM on an equally spaced grid. Since increments are stationary, we can efficiently simulate them using the circulant embedding method of Sect. 8.4.1. The next section motivates the use of increments and explains the types of dependencies present.

8.2.2 *Correlation and Long-Range Dependence of Increments*

To generate *fBM* on the equally spaced grid $0 = t_0 < t_1 < t_2 < \cdots < t_n = 1$, one starts with the increment process $Y_n = (B_n^H - B_{n-1}^H)$, $n \geq 1$. The time series Y_n can be characterized as a discrete stationary Gaussian sequence with zero mean and covariance function

$$Cov(Y_{n+k}, Y_n) = \frac{1}{2}\sigma^2((k+1)^{2H} + (k-1)^{2H} - 2k^{2H}) \tag{8.2}$$

This is called fractional Gaussian noise with Hurst parameter H. In particular, for $H = \frac{1}{2}$, the increments are independent and B^H becomes a Brownian motion.

The autocorrelation function for increments is obtained from (8.2) as

$$\rho_H(k) = \frac{1}{2}((k+1)^{2H} + (k-1)^{2H} - 2K^{2H}) \approx H(2H-1)k^{2(H-1)}, \tag{8.3}$$

where the approximation holds as k increases. The parameter H controls the regularity of trajectories. For $\frac{1}{2} < H < 1$, the autocorrelations $\rho_H(k) > 0$ are positive and decay slowly to zero though $\Sigma_{n=1}^{\infty}\rho_H(n) = \infty$, i.e. they exhibit long-range dependence. For $0 < H < \frac{1}{2}$, the autocorrelations $\rho_H(k) < 0$ are negative and decay to zero with a rate faster than $\frac{1}{n}$ and hence, the increments demonstrate short-range dependence, i.e. $\Sigma_{n=1}^{\infty}\rho_H(n) < \infty$.

8.3 Fractional Ornstein–Uhlenbeck Process of the Second Kind

The fractional Ornstein–Uhlenbeck process of the second kind (fOU_2) represents a class of stationary processes and was first introduced by Kaarakka and Salminen [13] as a way to extend Ornstein–Uhlenbeck diffusion. The new process is defined via Lamperti transformation of a fractional Brownian motion. The stochastic integrals below should be interpreted as Riemann–Stieltjes integrals, see a recent dissertation by Azmoodeh [3] for the detailed account of the matter.

The initial point in the construction of fOU_2 is an Ornstein–Uhlenbeck (OU) process [16], also widely known as the Vasicek model [17]. It is a unique strong solution of the stochastic differential equation

$$dU_t = \theta(\mu - U_t)dt + \sigma dB_t, \tag{8.4}$$

where $B = B_{t\geq 0}$ is a Brownian motion. The parameter μ is interpreted as the long-run equilibrium value of the process, σ is the volatility, and θ is the speed of reversion, i.e. the process oscillates around some equilibrium value.

The fractional Ornstein–Uhlenbeck (fOU) process X_t can be obtained as a solution of the following stochastic differential equation

$$dX_t = \theta(\mu - X_t)dt + \sigma dB_t^H, \tag{8.5}$$

where $\mu, \sigma, \theta > 0$ are parameters and $B^H = B_t^H$ as $t \geq 0$ is a fractional Brownian motion with Hurst parameter $\frac{1}{2} < H < 1$. It was named fractional Ornstein–Uhlenbeck process of the first kind (fOU_1) by Kaarakka and Salminen [13].

By applying Lamperti transformation to fractional Brownian motion B^H, let us define the following process $Y_t = \int_0^t e^{-s} dB_{a_s}^H$, where $a_s = He^{\frac{s}{H}}$. Substituting Y_t for fBM in (8.5), the following stochastic differential equation

$$dX_t = \theta(\mu - X_t)dt + \sigma dY_t, \tag{8.6}$$

has a solution that can be obtained by applying $Itô's$ lemma to $e^{\theta t}X_t$ and can be represented in the following integral form

$$X_t = e^{-\theta t}X_0 + \mu(1 - e^{-\theta t}) + \sigma \int_0^t e^{\theta(s-t)} dY_s, \tag{8.7}$$

where the last term is understood as a path-wise Riemann–Stiltjes integral. Following the terminology of Kaarakka and Salminen [13], we call this fractional Ornstein–Uhlenbeck process of the second kind (fOU_2).

It can be characterized by the following covariance function:

$$\gamma_U(H, \theta, t) = H(2H-1)e^{-\theta t} \int_{-\infty}^{t} \int_{-\infty}^{0} \frac{e^{(\theta - 1 + \frac{1}{H})(u+v)}}{|e^{\frac{u}{H}} - e^{\frac{v}{H}}|^{2(1-H)}} du dv, \tag{8.8}$$

where θ is the drift parameter from Eq. (8.6), see Andreev and Morlanes [1] for a detailed discussion on how to calculate the double integral. In particular, $Var(X_t) = \frac{1}{\theta} C(H, \theta) e^{-\theta t} \cdot B(H(\theta - 1) + 1, 2H - 1)$. In contrast to fBM, fOU_2 is short-range-dependent for all values of H, i.e. $\int_{-\infty}^{\infty} \gamma_U(H, \theta, u)du = 0$ for all H and θ. Furthermore, for $H = \frac{1}{2}$, the covariance function is zero and the process coincides with Ornstein–Uhlenbeck diffusion, see Kaarakka and Salminen [13]. Following the interpretation of the stochastic integral as a path-wise Riemann–Stiltjes integral, it can be shown that other properties of the fOU_2 are similar to those of fractional Brownian motion.

Below is a short sketch on how to discretize the double integral in (8.8) and reduce it to a single integral. For an extensive derivation and detailed discussion on the range of parameters where the procedure is applicable, we refer to Andreev and Morlanes [1]. We set $t = t_k = k\Delta t$ and $s = 0$ to show the step of the grid. This representation is crucial in order to perform the suggested simulation scheme.

$$
\begin{aligned}
&\int_{-\infty}^{t}\int_{-\infty}^{s} H^{2(H-1)}\frac{e^{(\theta-1+\frac{1}{H})(u+v)}}{|e^{\frac{u}{H}}-e^{\frac{v}{H}}|^{2(1-H)}}dudv\\
&=\int_0^{a_t}\int_0^{a_s}(mn)^{(\theta-1)H}|m-n|^{2(H-1)}dmdn\\
&=\int_0^{a_s}\int_0^{a_s}(mn)^{(\theta-1)H}|m-n|^{2(H-1)}dmdn+\\
&+\int_{a_s}^{a_t}\int_0^{a_s}(mn)^{(\theta-1)H}|m-n|^{2(H-1)}dmdn\\
&\approx\int_0^{a_s}m^{2\theta H-1}\int_0^1\xi^{(\theta-1)H}|1-\xi|^{2(H-1)}d\xi dm+\\
&+\int_{a_s}^{a_t}m^{2\theta H-1}\int_0^{a_s/m}\xi^{(\theta-1)H}|1-\xi|^{2(H-1)}d\xi dm \qquad (8.9)\\
&=\frac{a_s^{2\theta H}}{\theta H}B((\theta-1)H+1,2H-1))+\\
&+\int_{a_s}^{a_t}m^{2\theta H-1}B(a_s/m;(\theta-1)H+1,2H-1)dm\\
&=\frac{1}{\theta}B((\theta-1)H+1,2H-1))+\\
&+\int_H^{He^{\frac{k}{H}}}n^{2\theta H-1}B\left(\frac{H}{n};(\theta-1)H+1,2H-1\right)dn
\end{aligned}
$$

where $\mathbf{B}(\cdot;\cdot,\cdot)$ and $\mathbf{B}(\cdot,\cdot)$ denote the incomplete beta and the beta functions respectively. We use this discrete representation in the next section in order to build an efficient algorithm to simulate from fOU_2 by capitalizing on circulant property that allows to reduce complexity of simulations.

8.4 Simulation of fOU_2 Using *CEM*

The *Circulant Embedding Method* (CEM) is one of the prime choices to simulate from stationary Gaussian processes and we adapt it for simulation from fractional Ornstein–Uhlenbeck process of the second kind. The major difficulty in application to fOU_2 is to build an equally spaced grid for the covariance function (8.8).

We start with a standard equally spaced fine grid $\Delta t = T/N$ of interval $[0, T]$, where N is a large integer and $t_j = j\Delta t$, $X_{t_j} = X_j$. Next, the covariance function given by (8.8) needs to be presented as a function of the increment $k = t_i - t_j = |i - j| * \Delta t$

$$R(H, \theta, k) = C(H, \theta)e^{-\theta k}\left(\frac{1}{\theta}B((\theta-1)H+1, 2H-1)+\right.$$
$$\left.+\int_H^{He^{\frac{k}{H}}} n^{2\theta H-1}B\left(\frac{H}{n}; (\theta-1)H+1, 2H-1\right)dn\right), \tag{8.10}$$

for $k = 0, 1, \ldots, N-1$. As a preparatory step, we modify fOU_2 in order to be able to normalize the discretized covariance function as follows

$$\tilde{X}_j = \beta(H, \theta)^{-\frac{1}{2}} X_j, \tag{8.11}$$

where $\beta(H, \theta) = H^{2H(1-\theta)}H(2H-1)\cdot\theta^{-1}\cdot B(H(\theta-1)+1, 2H-1)$. The new process has the following covariance function

$$\tilde{R}(H, \theta, k) = \beta(H, \theta)^{-1}\cdot R(H, \theta, k),\ k = 0, 1, \ldots, N-1 \tag{8.12}$$

This is the setting that is suitable for utilization of *CEM* to generate from the normalized fOU_2 process $\tilde{X}$. The step-by-step procedure is explained in the next section and then summarized by Algorithm 1. Finally, one needs to apply the following transformation

$$X_j = \sqrt{\beta(H, \theta)}\cdot\tilde{X}_j \tag{8.13}$$

in order to recover the original fOU_2.

8.4.1 Circulant Embedding Method in a Nutshell

Algorithm 1 summarizes the step-by-step procedure to simulate from the process $\tilde{X}_j$, defined as the normalized fOU_2 in Eq. (8.11). It becomes possible due to a fact that the covariance matrix of a stationary discrete Gaussian process can be embedded into a so-called circulant matrix. This latter matrix should be non-negative definite for the algorithm to work. The implementation for the fOU_2 is given below.

The advantage of *CEM* to other available methods is that the circulant matrix can be diagonalized explicitly, and computations be performed efficiently using the Fast Fourier Transform (FFT) algorithm. We embed the covariance matrix of $\tilde{X}_j$ into a circulant matrix and describe the procedure of embedding using an illustrative example in the next section.

The circulant matrix $C \in R^{N\times N}$ with first column c_1 allows for decomposition $C = WDW^*$, where W is a Fourier matrix and D is a diagonal matrix with diagonal $\lambda = \sqrt{N} \cdot W \cdot c_1$. The columns of W are the eigenvectors of C and D contains the eigenvalues. If all the eigenvalues are non-negative, then by defining $R := WD^{1/2}$, C can be factorized as $C = RR^*$.

Next, one generates a complex value vector $\hat{\mathbf{X}} = \mathrm{R}\boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is a complex Gaussian vector of length N and distribution $\boldsymbol{\xi} \sim CN(0, 2I_N))$ so that $\hat{\boldsymbol{X}} \sim CN(\mathbf{0}, 2C)$. From the real and the imaginary part of vector $\hat{\boldsymbol{X}} = \hat{\boldsymbol{X}}_{\mathbf{1}} + i\hat{\boldsymbol{X}}_{\mathbf{2}}$, we obtain two sequences of length N with the covariance matrix equal to the circulant matrix C, i.e. $\hat{\boldsymbol{X}}_1, \hat{\boldsymbol{X}}_{\mathbf{2}} \sim N(0, C)$.

We finally extract two normalized fOU_2 vectors $\tilde{\boldsymbol{X}}_{\mathbf{1}}$ and $\tilde{\boldsymbol{X}}_{\mathbf{2}}$ with covariance matrix entries given by $R(\tilde{H}, \theta, k)$ from $\hat{\boldsymbol{X}}_1$ and $\hat{\boldsymbol{X}}_{\mathbf{2}}$.

Algorithm 1 Generation of two fOU_2 vectors of length N using *CEM*

1: Consider the $0 < t_0 < t_1 < \cdots < t_N = T$ grid
2: Generate the initial value X_0
3: Embed the covariance matrix in a circular matrix C
4: Factorize $C = RR^*$ by fast Fourier transform or using matrix properties
5: Generate a complex vector $\boldsymbol{\xi} = \boldsymbol{\xi}_{\mathbf{1}} + i\boldsymbol{\xi}_{\mathbf{2}}$ where $\boldsymbol{\xi}_{\mathbf{1}}, \boldsymbol{\xi}_{\mathbf{2}} \sim N(0, I_N), iid$
6: Evaluate $\boldsymbol{X} = R\boldsymbol{\xi}$ by fast Fourier transform
7: Get the real part $R(\boldsymbol{X})$ and the imaginary part $I(\boldsymbol{X})$
8: Save the first N values of the $R(\boldsymbol{X})$ and $I(\boldsymbol{X})$

8.4.2 Illustrative Example

We illustrate Algorithm 1 by providing a step-by-step procedure on how to simulate fOU_2 vector $\boldsymbol{X} = (X_0, X_1, X_2, X_3, X_4)$ with parameters $H = 0.8, \theta = 3$ on interval $[0, 1]$ with N=*5*. Straightforward calculations show that the normalized fOU_2 vector $\tilde{\mathbf{X}}$ has scaling factor $\beta(0.8, 3) = 0.2870$ and the following covariance matrix, up to rounding error in the second digit:

$$\begin{pmatrix} 1 & 0.25 & 0.15 & 0.11 & 0.09 \\ 0.25 & 1 & 0.25 & 0.15 & 0.11 \\ 0.15 & 0.25 & 1 & 0.25 & 0.15 \\ 0.11 & 0.15 & 0.25 & 1 & 0.25 \\ 0.09 & 0.11 & 0.15 & 0.25 & 1 \end{pmatrix}$$

which is a symmetric Toeplitz matrix, not necessarily a circulant matrix. This matrix can be embedded into a larger symmetric circulant matrix as follows

$$\left(\begin{array}{cccccccc} 1 & 0.25 & 0.15 & 0.11 & 0.09 & 0.11 & 0.15 & 0.25 \\ 0.25 & 1 & 0.25 & 0.15 & 0.11 & 0.09 & 0.11 & 0.15 \\ 0.15 & 0.25 & 1 & 0.25 & 0.15 & 0.11 & 0.09 & 0.11 \\ 0.11 & 0.15 & 0.25 & 1 & 0.25 & 0.15 & 0.11 & 0.09 \\ 0.09 & 0.11 & 0.15 & 0.25 & 1 & 0.25 & 0.15 & 0.11 \\ \hline 0.11 & 0.09 & 0.11 & 0.15 & 0.25 & 1 & 0.25 & 0.15 \\ 0.15 & 0.11 & 0.09 & 0.11 & 0.15 & 0.25 & 1 & 0.25 \\ 0.25 & 0.15 & 0.11 & 0.09 & 0.11 & 0.15 & 0.25 & 1 \end{array}\right)$$

where the first column is $c_1 = (1, 0.25, 0.15, 0.11, 0.09, 0.11, 0.15, 0.25)^T$ which allows for a decomposition of the following form: $\frac{1}{\sqrt{8}} W D W^*$, where

$$W = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega & -i & -i\omega & -1 & -\omega & i & i\omega \\ 1 & -i & -1 & i & 1 & -i & -1 & i \\ 1 & -i\omega & i & \omega & -1 & i\omega & -i & -\omega \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -\omega & -i & i\omega & -1 & \omega & i & -i\omega \\ 1 & i & -1 & -i & 1 & i & -1 & -i \\ 1 & i\omega & i & -\omega & -1 & -i\omega & -i & \omega \end{pmatrix}$$

$$D = \begin{pmatrix} 2.12 & & & & & & & \\ & 1.11 & & & & & & \\ & & 1.11 & & & & & \\ & & & 0.66 & & & & \\ & & & & 0.79 & & & \\ & & & & & 0.79 & & \\ & & & & & & 0.72 & \\ & & & & & & & 0.72 \end{pmatrix}$$

and W^* is the conjugate transpose of W. All eigenvalues of the minimal circulant extension are non-negative, i.e. it defines a valid covariance matrix. We generate $\hat{\boldsymbol{X}} = W D^{1/2} \boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim CN(\mathbf{0}, 2I_8)$.

We finally consider the first five elements of the real part of $\hat{\boldsymbol{X}}$ and multiply them by $\sqrt{\beta(0.8, 3)}$, thus obtaining a fOU_2 sample vector $\mathbf{X}$ of length 5, as shown in Table 8.1.

8.5 Conclusion

In this paper, we show how to simulate one-dimensional fOU_2 by a circulant embedding method. The two main steps are to embed the covariance matrix into a circulant matrix and then to use a fast Fourier transform (EFT) algorithm. We exemplify the

Table 8.1 First five samples for fOU_2 with $H = 0.8$ and $\theta = 3$

Time	$\hat{X}$	$\tilde{X}_1$	X
0	0.45−i0.04	0.45	0.24
1	0.37+i0.50	0.37	0.20
2	0.04−i0.49	0.04	0.02
3	0.07−i0.27	0.07	0.04
4	−0.25−i0.20	−0.25	−0.13
5	−0.24+i0.44		
6	−0.04+i0.44		
7	−0.12−i0.09		

method with simulation of a random vector fOU_2 with parameters $H = 0.8$ and $\theta = 3$, and length $N = 5$.

The circulant embedding method allows to extend the method to generate two- and three-dimensional fOU_2 vectors. The fOU_2 covariance matrix is then embedded into a block circulant matrix with each block being circulant itself. Two- and three-dimensional FFT techniques are then applied. Finally, if we wish to simulate a non-Gaussian fOU_2 vector, the circulant embedding method can be combined with e.g a memoryless non-linear transformation, see Grigoriu [11].

Acknowledgements We thank Michael Carlson for assistance with final English check.

References

1. Andreev, A., Morlanes, J.I.: Simulation-based studies of covariance structure for fractional Ornstein-Uhlenbeck process of the second kind, submitted manuscript (2018)
2. Asmussen, S.: Stochastic simulation with a view towards stochastic processes. University of Aarhus, Center for Mathematical Physics and Stochastics (MaPhySto) [MPS] (1998)
3. Azmoodeh, E.: Riemann-Stiltjes integrals with respect to fractional Brownian motion and applications. Ph.D. dissertation, Helsinki University of Technology Institute of Mathematics (2010)
4. Azmoodeh, E., Morlanes, J.I.: Drift parameter estimation for fractional Ornstein-Uhlenbeck process of the second kind. Statistics **49**(1), 1–18 (2015)
5. Azmoodeh, E., Viitasaari, L.: Parameter estimation based on discrete observations of fractional Ornstein-Uhlenbeck process of the second kind. Stat. Inference Stoch. Process. **18**(3), 205–227 (2015)
6. Chan, G., Wood, A.T.A.: An algorithm for simulating stationary Gaussian random fields. J. R. Stat. Soc.: Ser. C (Appl. Stat.) **46**(1), 171–181 (1997)
7. Chan, G., Wood, A.T.A.: Simulation of stationary Gaussian vector fields. Stat. Comput. **9**(4), 265–268 (1999)
8. Davies, R.B., Harte, D.S.: Tests for Hurst effect. Biometrika **74**, 95–102 (1987)
9. Dietrich, C.R., Newsam, G.N.: Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. SIAM J. Sci. Stat. Comput. **18**(4), 1088–1107 (1997)
10. Gneiting, T., Sevcikova, H., Percival, D.B., Schalther, M., Jiang, Y.: Fast and exact simulation of large Gaussian lattice systems: exploring the limits. J. Comput. Graph. Stat. **15**(3) (2006)

11. Grigoriu, M.: Simulation of stationary non-Gaussian translation processes. J. Eng. Mech. **124**(2), 121–126 (1998)
12. Hosking, J.R.M.: Modeling persistence in hydrological time series using fractional differencing. Water Resour. Res. **20**(12), 1898–1908 (1984)
13. Kaarakka, T., Salminen, P.: On fractional Ornstein-Uhlenbeck processes. Commun. Stoch. Anal. **5**(1), 121–133 (2011)
14. Mandelbrot, B., van Ness, J.: Fractional Brownian motions, fractional noises and applications. SIAM Rev. **10**(4), 422–437 (1968)
15. Nualart, D.: The Malliavin Calculus and Related Topics, Probability and Its Applications. Springer, Berlin (2006)
16. Uhlenbeck, G.E., Ornstein, L.S.: On the theory of Brownian motion. Phys. Rev. **36**(5), 823 (1930)
17. Vasicek, O.: An equilibrium characterization of the term structure. J. Financ. Econ. **5**(2), 177–188 (1977)
18. Wood, A.T.A., Chan, G.: Simulation of stationary Gaussian processes in [0,1]. J. Comput. Graph. Stat. **3**(4), 409–432 (1994)

Chapter 9
Constructive Martingale Representation in Functional Itô Calculus: A Local Martingale Extension

Kristoffer Lindensjö

Abstract The constructive martingale representation theorem of functional Itô calculus is extended, from the space of square integrable martingales, to the space of local martingales. The setting is that of an augmented filtration generated by a Wiener process.

Keywords Functional Itô calculus · Martingale representation

9.1 Introduction

Consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which lives an n-dimensional Wiener process W. Let $\underline{\mathcal{F}} = (\mathcal{F}_t)_{0\leq t\leq T}$ denote the augmentation under $\mathbb{P}$ of the filtration generated by W until the constant terminal time $T < \infty$. One of the main results of Itô calculus is the martingale representation theorem which in the present setting is as follows: *Let M be a RCLL local martingale relative to $(\mathbb{P}, \underline{\mathcal{F}})$, then there exists a progressively measurable n-dimensional process φ such that*

$$M(t) = M(0) + \int_0^t \varphi(s)' dW(s),\ 0 \leq t \leq T, \text{ and } \int_0^T |\varphi(t)|^2 dt < \infty \text{ a.s.}$$

In particular, M has continuous sample paths a.s.

Considerable effort has in the literature been made in order to find explicit formulas for the integrand φ, i.e. in order to find constructive representations of martingales, mainly using Malliavin calculus, see e.g. [8, 15, 16, 20] and the references therein. The recently developed *functional Itô calculus* includes a new type of constructive representation of square integrable martingales due to Cont and Fournié see e.g. [1, 3–5]. The main result of the present paper is an extension of this result to local martingales.

K. Lindensjö (✉)
Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden
e-mail: kristoffer.lindensjo@math.su.se

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_9

The organization of the paper is as follows. Section 9.2 is based on [1] and contains a brief and heuristic account of the relevant parts of functional Itô calculus including the constructive martingale representation theorem for square integrable martingales. Section 9.3 contains the local martingale extension of this theorem and a simple example.

9.2 Constructive Representation of Square Integrable Martingales

Denote an n-dimensional sample path by ω. Denote a sample path stopped at t by ω_t, i.e. let $\omega_t(s) = \omega(t \wedge s), 0 \leq s \leq T$. Consider a real-valued functional of sample paths $F(t, \omega)$ which is *non-anticipative* (essentially meaning that $F(t, \omega) = F(t, \omega_t)$). The *horizontal derivative* at (t, ω) is defined by

$$\mathcal{D}F(t, \omega) = \lim_{h \searrow 0} \frac{F(t + h, \omega_t) - F(t, \omega_t)}{h}.$$

The *vertical derivative* at (t, ω) is defined by

$$\nabla_\omega F(t, \omega) = (\partial_i F(t, \omega), i = 1, ..., n)',$$

where

$$\partial_i F(t, \omega) = \lim_{h \to 0} \frac{F(t, \omega_t + h e_i I_{[t,T]}) - F(t, \omega_t)}{h}.$$

Higher order vertical derivatives are obtained by vertically differentiating vertical derivatives.

One of the main results of functional Itô calculus is the functional Itô formula, which is just the standard Itô formula with the usual time and space derivatives replaced by the horizontal and vertical derivatives. If the functional F is sufficiently regular (regarding e.g. continuity and boundedness of its derivatives), which we write as $F \in \mathbb{C}_b^{1,2}$, then the functional Itô formula holds, see [1, ch. 5,6]. We remark that [12] contains another version of this result.

Using the functional Itô formula it easy to see that if Z is a martingale satisfying

$$Z(t) = F(t, W_t) \;\; dt \times d\mathbb{P}\text{-a.e., with } F \in \mathbb{C}_b^{1,2}, \tag{9.1}$$

then, for every $t \in [0, T]$,

$$Z(t) = Z(0) + \int_0^t \nabla_\omega F(s, W_s)' dW(s) \;\; a.s.$$

We may therefore define the vertical derivative with respect to the process W of a martingale Z satisfying (9.1) as the $dt \times d\mathbb{P}$-a.e. unique process $\nabla_W Z$ given by

$$\nabla_W Z(t) = \nabla_\omega F(t, W_t), 0 \leq t \leq T. \tag{9.2}$$

Let $\mathcal{C}_b^{1,2}(W)$ be the space of processes Z which allow the representation in (9.1). Let $\mathcal{L}^2(W)$ be the space of progressively measurable processes φ satisfying the condition $E[\int_0^T \varphi(s)'\varphi(s)ds] < \infty$. Let $\mathcal{M}^2(W)$ be the space of square integrable martingales with initial value 0. Let $D(W) = \mathcal{C}_b^{1,2}(W) \cap \mathcal{M}^2(W)$.

It can be shown that $\{\nabla_W Z : Z \in D(W)\}$ is dense in $\mathcal{L}^2(W)$ and that $D(W)$ is dense in $\mathcal{M}^2(W)$ [1, ch. 7]. Using this it is possible to show that the vertical derivative operator $\nabla_W(\cdot)$ admits a unique extension to $\mathcal{M}^2(W)$, in the following sense: *For $Y \in \mathcal{M}^2(W)$ the (weak) vertical derivative $\nabla_W Y$ is the unique element in $\mathcal{L}^2(W)$ satisfying*

$$E[Y(T)Z(T)] = E\left[\int_0^T \nabla_W Y(t)'\nabla_W Z(t)dt\right] \tag{9.3}$$

for every $Z \in D(W)$, where $\nabla_W Z$ is defined in (9.2). The constructive martingale representation theorem ([1, ch. 7]) follows:

Theorem 9.1 (Cont and Fournié) *For any square integrable martingale Y relative to $(\mathbb{P}, \underline{\mathcal{F}})$ and every $t \in [0, T]$,*

$$Y(t) = Y(0) + \int_0^t \nabla_W Y(s)'dW(s) \quad a.s.$$

9.3 Constructive Representation of Local Martingales

This section contains an extension of the vertical derivative $\nabla_W(\cdot)$ and the constructive martingale representation in Theorem 9.1 to local martingales. Let $\mathcal{M}^{\text{loc}}(W)$ denote the space of local martingales relative to $(\mathbb{P}, \underline{\mathcal{F}})$ with initial value zero and RCLL sample paths. In Theorem 9.2 we extend the vertical derivative to $\mathcal{M}^{\text{loc}}(W)$. Using this extension we can formulate the constructive martingale representation theorem also for local martingales, see Theorem 9.3.

Before extending the definition of the vertical derivative to $\mathcal{M}^{\text{loc}}(W)$ we recall the definition of a local martingale.

Definition 9.1 M is said to be a local martingale if there exists a sequence of non-decreasing stopping times $\{\theta_n\}$ with $\lim_{n\to\infty} \theta_n = \infty$ a.s. such that the stopped local martingale $M(\cdot \wedge \theta_n)$ is a martingale for each $n \geq 1$.

Theorem 9.2 (Definition of $\nabla_W(\cdot)$ on $\mathcal{M}^{loc}(W)$)

- *There exists a progressively measurable $dt \times d\mathbb{P}$-a.e. unique extension of the vertical derivative $\nabla_W(\cdot)$ from $\mathcal{M}^2(W)$ to $\mathcal{M}^{loc}(W)$, such that, for $M \in \mathcal{M}^{loc}(W)$,*

$$\begin{array}{c} M(t) = \int_0^t \nabla_W M(s)' dW(s), 0 \le t \le T, \text{ and} \\ \int_0^T |\nabla_W M(t)|^2 dt < \infty a.s. \end{array} \tag{9.4}$$

- *Specifically, for $M \in \mathcal{M}^{loc}(W)$ the vertical derivative $\nabla_W M$ is defined as the progressively measurable $dt \times d\mathbb{P}$-a.e. unique process satisfying*

$$\nabla_W M(t) = \lim_{n\to\infty} \nabla_W M_n(t) \;\; dt \times d\mathbb{P}\text{-a.e.} \tag{9.5}$$

where $\nabla_W M_n$ is the vertical derivative of $M_n := M(\cdot \wedge \tau_n) \in \mathcal{M}^2(W)$ and τ_n is given by

$$\tau_n = \theta_n \wedge \inf\{s \in [0,T] : |M(s)| \ge n\} \wedge T \tag{9.6}$$

where $\{\theta_n\}$ is an arbitrary sequence of stopping times of the kind described in Definition 9.1.

Remark 9.1 Note that if M in Theorem 9.2 satisfies

$$M(t) = \int_0^t \gamma(s)' dW(s), 0 \le t \le T \text{ a.s.}$$

for some process γ, then $\gamma = \nabla_W M \; dt \times d\mathbb{P}$-a.e. It follows that the extended vertical derivative $\nabla_W M$ defined in Theorem 9.2 does not depend (modulo possibly on a null set $dt \times d\mathbb{P}$) on the particulars of the chosen stopping times $\{\theta_n\}$.

Proof The martingale representation theorem implies that, for $M \in \mathcal{M}^{\text{loc}}(W)$, there exists a progressively measurable process φ satisfying

$$M(t) = \int_0^t \varphi(s)' dW(s), 0 \le t \le T, \text{ and } \int_0^T |\varphi(t)|^2 dt < \infty \text{ a.s.} \tag{9.7}$$

Therefore, if we can prove that

$$\lim_{n\to\infty} \nabla_W M_n(t) = \varphi(t) \;\; dt \times d\mathbb{P}\text{-a.e.}, \tag{9.8}$$

then it follows that there exists a progressively measurable process, denote it by $\nabla_W M$, which is $dt \times d\mathbb{P}$-a.e. uniquely defined by (9.5) and satisfies

$$\nabla_W M(t) = \varphi(t) \;\; dt \times d\mathbb{P}\text{-a.e.},$$

which in turn implies that the integrals of $\nabla_W M$ and φ coincide in the way that (9.7) implies (9.4). All we have to do is therefore to prove that (9.8) holds.

Let us recall some results about stopping times and martingales. The stopped local martingale $M(\cdot \wedge \theta_n)$ is a martingale for each n, by Definition 9.1. Stopped RCLL martingales are martingales. The minimum of two stopping times is a stopping time and the hitting time

$$\inf\{s \in [0, T] : |M(s)| \geq n\}$$

is, for each n, in the present setting, a stopping time. Using these results we obtain that $M(\cdot \wedge \theta_n \wedge \inf\{s \in [0, T] : |M(s)| \geq n\} \wedge T) = M(\cdot \wedge \tau_n)$ is a martingale, for each n. Moreover, M is by the standard martingale representation result a.s. continuous. Hence, we may define a sequence of, a.s. continuous, martingales $\{M_n\}$ by

$$M_n = M(\cdot \wedge \tau_n) = \int_0^{\cdot \wedge \tau_n} \varphi(s)' dW(s) \text{ a.s.} \tag{9.9}$$

where the last equality follows from (9.7). Now, use the definition of τ_n in (9.6) to see that

$$|M_n(t)| = \left| \int_0^{t \wedge \tau_n} \varphi(s)' dW(s) \right| \leq n \text{ a.s.}$$

for any t and n, and that in particular M_n is, for each n, a square integrable martingale. Moreover, (9.9) implies that M_n satisfies

$$M_n(t) = \int_0^t I_{\{s \leq \tau_n\}} \varphi(s)' dW(s), 0 \leq t \leq T \text{ a.s.} \tag{9.10}$$

Since each M_n is a square integrable martingale we may use Theorem 9.1 on M_n, which together with (9.10) implies that

$$\begin{aligned} M_n(t) &= \int_0^t \nabla_W M_n(s)' dW(s) \\ &= \int_0^t I_{\{s \leq \tau_n\}} \varphi(s)' dW(s), 0 \leq t \leq T \text{ a.s.} \end{aligned} \tag{9.11}$$

where $\nabla_W M_n$ is the vertical derivative of M_n with respect to W (defined in (9.3)) and where we also used the continuity of the Itô integrals. The equality of the two Itô integrals in (9.11) implies that

$$\nabla_W M_n(t) = I_{\{t \leq \tau_n\}} \varphi(t) \ \ dt \times d\mathbb{P}\text{-a.e.} \tag{9.12}$$

The local martingale property of M implies that $\lim_{n \to \infty} \theta_n = \infty \ a.s.$ Using this and the definition of τ_n in (9.6) we conclude that for almost every $\omega \in \Omega$ and each

$t \in [0, T]$ there exists an $N(\omega, t)$ such that

$$n \geq N(\omega, t) \Rightarrow \sup_{0 \leq s \leq t} |M(\omega, s)| \leq n \text{ and } t \leq \theta_n(\omega) \Rightarrow t \leq \tau_n(\omega). \quad (9.13)$$

It follows from (9.12) and (9.13) that there exists an $N(\omega, t)$ such that

$$n \geq N(\omega, t) \Rightarrow \nabla_W M_n(\omega, t) = \varphi(\omega, t) \ \ dt \times d\mathbb{P}\text{-a.e.}$$

which means that (9.8) holds. □

If M is a RCLL local martingale then $M - M(0) \in \mathcal{M}^{\text{loc}}(W)$, which implies that $\nabla_W(M - M(0))$ is defined in Theorem 9.2. This observation allows us to extend the definition of the vertical derivative to RCLL local martingales not necessarily starting at zero in the following obvious way.

Definition 9.2 The vertical derivative of a local martingale M relative to $(\mathbb{P}, \underline{\mathcal{F}})$ with RCLL sample paths is defined as the progressively measurable $dt \times d\mathbb{P}$-a.e. unique process $\nabla_W M$ satisfying

$$\nabla_W M(t) = \nabla_W(M - M(0))(t), 0 \leq t \leq T, \quad (9.14)$$

where $\nabla_W(M - M(0))(t)$ is defined in Theorem 9.2.

The following result is an immediate consequence of Theorem 9.2 and Definition 9.2.

Theorem 9.3 *If M is a local martingale relative to $(\mathbb{P}, \underline{\mathcal{F}})$ with RCLL sample paths, then*

$$M(t) = M(0) + \int_0^t \nabla_W M(s)' dW(s), 0 \leq t \leq T, \text{ and}$$
$$\int_0^T |\nabla_W M(t)|^2 dt < \infty \ a.s.,$$

where $\nabla_W M(s)$ is defined in Definition 9.2.

Let us try to clarify the theory by studying a simple example. It is straightforward to extend the results above to the case when the Wiener process W is replaced by an adapted process X given by

$$X(t) = X(0) + \int_0^t \sigma(s) dW(s), \quad (9.15)$$

where σ is a matrix-valued adapted process satisfying suitable assumptions, mainly invertibility, see also [1, 4]. Thus, a local martingale M can be represented as

$$M(t) - M(0) = \int_0^t \nabla_W M(s)' dW(s) = \int_0^t \nabla_X M(s)' dX(s),$$

and the relationship between the vertical derivatives with respect to W and X is $\nabla_W M(t)' = (\nabla_X M(t)')\sigma(t)$, cf. (9.15). As example consider the one-dimensional case and let X with $X(0) = 0$ be given by (9.15) under the assumption that $\sigma(s)$ is a deterministic function of time and let M be given by $M(t) = F(t, X_t)$ where F is the non-anticipative functional $F(t, \omega) = \omega^3(t) - 3\int_0^t \omega(s)\sigma^2(s)ds$, i.e. let M be the local martingale defined by

$$M(t) = X^3(t) - 3\int_0^t X(s)\sigma^2(s)ds.$$

In this case the vertical derivative simplifies to the standard derivative, that is, $\nabla F_\omega(t, \omega) = 3\omega^2(t)$, see also [1, 4] (we remark that the horizontal derivative is $\mathcal{D}F(t, \omega) = -3\omega(t)\sigma^2(t)$). In this case, $\nabla_X M(t) = 3X^2(t)$ and

$$M(t) = \int_0^t 3X^2(s)dX(s) = \int_0^t 3X^2(s)\sigma(s)dW(s),$$

which we remark is easily found using the standard Itô formula. Note that this also means that $\nabla_W M(t) = 3X^2(t)\sigma(t) = \nabla_X M(t)\sigma(t)$.

Concluding Remarks

Many of the applications that rely on martingale representation are within mathematical finance. A particular application that may benefit from the local martingale extension of the present paper is optimal investment theory, in which the discounted (using the state price density) optimal wealth process is a (not necessarily square integrable) martingale, see e.g. [9, ch. 3], see also [13]. In particular, using functional Itô calculus it is possible to derive an explicit formula for the optimal portfolio in terms of the vertical derivative of the discounted optimal wealth process, see also [14]. Similar explicit formulas for optimal portfolios based on the Malliavin calculus approach to constructive martingale representation have, under restrictive assumptions, been studied extensively, see e.g. [2, 6, 7, 10, 11, 17–19]. The general connection between Malliavin calculus and functional Itô calculus is studied in e.g. [1, 4].

Acknowledgements The author is grateful to Mathias Lindholm for helpful discussions.

References

1. Bally, V., Caramellino, L., Cont, R., Utzet, F., Vives, J.: Stochastic Integration by Parts and Functional Itô Calculus. Springer, Berlin (2016)

2. Benth, F.E., Di Nunno, G., Løkka, A., Øksendal, B., Proske, F.: Explicit representation of the minimal variance portfolio in markets driven by Lévy processes. Math. Financ. **13**(1), 55–72 (2003)
3. Cont, R., Fournié, D.-A.: Change of variable formulas for non-anticipative functionals on path space. J. Funct. Anal. **259**(4), 1043–1072 (2010)
4. Cont, R., Fournié, D.-A.: Functional Itô calculus and stochastic integral representation of martingales. Ann. Prob. **41**(1), 109–133 (2013)
5. Cont, R., Lu, Y.: Weak approximation of martingale representations. Stoch. Process. Appl. **126**(3), 857–882 (2016)
6. Detemple, J., Rindisbacher, M.: Closed-form solutions for optimal portfolio selection with stochastic interest rate and investment constraints. Math. Financ. **15**(4), 539–568 (2005)
7. Nunno, Di G., Øksendal, B.: Optimal portfolio, partial information and Malliavin calculus. Stoch.: Int. J. Probab. Stoch. Process. **81**(3–4), 303–322 (2009)
8. Karatzas, I., Ocone, D.L., Li, J.: An extension of Clark's formula. Stoch.: Int. J. Probab. Stoch. Process. **37**(3), 127–131 (1991)
9. Karatzas, I., Shreve, S.E.: Methods of Mathematical Finance (Stochastic Modelling and Applied Probability). Springer, Berlin (1998)
10. Lakner, P.: Optimal trading strategy for an investor: the case of partial information. Stoch. Process. Appl. **76**(1), 77–97 (1998)
11. Lakner, P., Nygren, L.M.: Portfolio optimization with downside constraints. Math. Financ. **16**(2), 283–299 (2006)
12. Levental, S., Schroder, M., Sinha, S.: A simple proof of functional Itô's lemma for semimartingales with an application. Stat. Probab. Lett. **83**(9), 2019–2026 (2013)
13. Lindensjö, K.: Optimal investment and consumption under partial information. Math. Meth. Oper. Res. **83**(1), 87–107 (2016)
14. Lindensjö, K.: An Explicit Formula for Optimal Portfolios in Complete Wiener Driven Markets: a Functional Itô Calculus Approach (2017). arXiv:1610.05018
15. Malliavin, P.: Stochastic Analysis, Grundlehren der Mathematischen Wissenschaften, vol. 313, pp XII, 347. Springer, Berlin (2015)
16. Nualart, D.: The Malliavin Calculus and Related Topics (Probability and Its Applications, 2nd edn. Springer, Berlin (2006)
17. Ocone, D.L., Karatzas, I.: A generalized Clark representation formula, with application to optimal portfolios. Stoch.: Int. J. Probab. Stoch. Process. **34**(3–4), 187–220 (1991)
18. Okur, Y.Y.: White noise generalization of the Clark-Ocone formula under change of measure. Stoch. Anal. Appl. **28**(6), 1106–1121 (2010)
19. Pham, H., Quenez, M.-C.: Optimal portfolio in partially observed stochastic volatility models. Ann. Appl. Probab. **11**(1), 210–238 (2001)
20. Rogers, L.C.G., Williams, D.: Diffusions, Markov Processes and Martingales, vol. 2, Itô calculus. Cambridge University Press, Cambridge (2000)

Chapter 10
Random Fields Related to the Symmetry Classes of Second-Order Symmetric Tensors

Anatoliy Malyarenko and Martin Ostoja-Starzewski

Abstract Under the change of basis in the three-dimensional space by means of an orthogonal matrix g, a matrix A of a linear operator is transformed as $A \mapsto gAg^{-1}$. Mathematically, the stationary subgroup of a *symmetric* matrix under the above action can be either $D_2 \times Z_2^c$, when all three eigenvalues of A are different, or $\mathrm{O}(2) \times Z_2^c$, when two of them are equal, or $\mathrm{O}(3)$, when all three eigenvalues are equal. Physically, one typical application relates to dependent quantities like a second-order symmetric stress (or strain) tensor. Another physical setting is that of dependent fields, such as conductivity with such three cases is the conductivity (or, similarly, permittivity, or anti-plane elasticity) second-rank tensor, which can be either orthotropic, transversely isotropic, or isotropic. For each of the above symmetry classes, we consider a homogeneous random field taking values in the fixed point set of the class that is invariant with respect to the natural representation of a certain closed subgroup of the orthogonal group. Such fields may model stochastic heat conduction, electric permittivity, etc. We find the spectral expansions of the introduced random fields.

Keywords Random field · Symmetry class · Spectral expansion

A. Malyarenko (✉)
Division of Applied Mathematics, School of Education Culture and Communication, Mälardalen University, 883, 72123 Västerås, Sweden
e-mail: anatoliy.malyarenko@mdh.se

M. Ostoja-Starzewski
University of Illinois at Urbana-Champaign, Urbana, IL 61801-2906, USA
e-mail: martinos@illinois.edu

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_10

10.1 Introduction

10.1.1 Why Tensor Random Fields?

The starting point for deterministic theories of continuum physics is the *field equation*

$$\mathcal{L}\vec{q} = \vec{f} \tag{10.1}$$

defined for a body $\mathcal{B}$ on some subset $\mathcal{D}$ of the d-dimensional Euclidean space $\mathbb{R}^d$, where $\mathcal{L}$ is a linear differential operator, $\vec{f}$ is a *source* or *forcing function*, and $\vec{q}$ is a solution field. This needs to be accompanied by appropriate boundary and/or initial conditions. A field theory is stochastic if either the operator $\mathcal{L}$ is random, or the forcing is random, or the boundary/initial conditions are random. In this paper we focus on the first case, so that (10.1) becomes $\mathcal{L}(\omega)\vec{q} = \vec{f}$, where, in the vein of random processes and fields, the randomness is indicated by the dependence of $\mathcal{L}(\omega)$ on an elementary event ω. Since the operator $\mathcal{L}$ is linear, it is usually described by a tensor-valued random field modelling, say, thermal conductivity, T, in a random medium.

An example of such a *stochastic boundary value problem* is to find a random field $T : \mathcal{D} \times \Omega \to \mathbb{R}$ such that

$$\nabla \cdot (\vec{C}(\vec{x}, \omega) \cdot \nabla T) = f(\vec{x}, \omega), \vec{x} \in \mathcal{D}, \;\; T(\vec{x}, \omega) = g(\vec{x}), \vec{x} \in \partial\mathcal{D}.$$

Several well-known analogs of stochastic conductivity problems described by elliptic equations of this type are given in Table 10.1.

Table 10.1 A collection of diverse physical problems governed by an elliptic-type equation with a random field of second-rank property, such as thermal conductivity. The anti-plane elasticity and torsion involve random fields on $\mathbb{R}^2$ only

Physical subject	T	∇T	$\vec{C}$	$\vec{q}$
Heat conduction	Temperature	Thermal gradient	Thermal conductivity	Heat flux
Anti-plane elasticity	Displacement	Strain	Elastic moduli	Stress
Torsion	Stress function	Strain	Shear moduli	Stress
Electrical conduction	Potential	Intensity	Electrical conductivity	Current density
Electrostatics	Potential	Intensity	Permittivity	Electric induction
Magnetostatics	Potential	Intensity	Magnetic permeability	Magnetic induction
Fickian diffusion	Concentration	Gradient	Diffusivity	Flux

Here f is the scalar field of a temperature source/sink, and $\vec{q}$ is the boundary value of T. While the above corresponds, in general, to property fields, another physical application of tensor-valued random fields applies to dependent fields such as stress, strain, etc. All the rank 2 tensor random fields in Table 10.1 have to be positive-definite point-wise.

10.1.2 Basic Concepts of Random Fields

Let $(\Omega, \mathfrak{F}, \mathsf{P})$ be a probability space, let $d \geq 2$ and $r \geq 0$ be integers, and let V be a subspace of the rth tensor power $(\mathbb{R}^d)^{\otimes r}$ with the convention $(\mathbb{R}^d)^{\otimes 0} = \mathbb{R}^1$. A *tensor-valued random field* is a function $\mathsf{T}(\vec{x}, \omega)\colon \mathbb{R}^d \times \Omega \to \mathsf{V}$ such that for any fixed $\vec{x}_0 \in \mathbb{R}^d$ the function $\mathsf{T}(\vec{x}_0, \omega)$ is a V-valued random tensor. We are interested in the case of $d = 3$ that corresponds to *space problems of continuum physics*.

Let $(\cdot, \cdot)$ be the restriction of the standard inner product in the space $(\mathbb{R}^d)^{\otimes r}$ to its subspace V, and let $\| \cdot \|$ be the corresponding norm. A random field $\mathsf{T}(\vec{x})$ is *second-order* if for any $\vec{x} \in \mathbb{R}^d$ we have $\mathsf{E}[\|\mathsf{T}(\vec{x})\|^2] < \infty$. A second-order random field $\mathsf{T}(\vec{x})$ is called *mean-square continuous* if for any $\vec{x}_0 \in \mathbb{R}^d$ we have $\lim_{\|\vec{x}-\vec{x}_0\|\to 0} \mathsf{E}[\|\mathsf{T}(\vec{x}) - \mathsf{T}(\vec{x}_0)\|^2] = 0$. In what follows we consider only mean-square continuous random fields.

Let $\langle \mathsf{T}(\vec{x}) \rangle = \mathsf{E}[\mathsf{T}(\vec{x})]$ be the *one-point correlation tensor* of the random field $\mathsf{T}(\vec{x})$, and let $\langle \mathsf{T}(\vec{x}), \mathsf{T}(\vec{y}) \rangle = \mathsf{E}[(\mathsf{T}(\vec{x}) - \langle \mathsf{T}(\vec{x}) \rangle) \otimes (\mathsf{T}(\vec{y}) - \langle \mathsf{T}(\vec{y}) \rangle)]$ be its *two-point correlation tensor*. A random field $\mathsf{T}(\vec{x})$ is called *wide-sense homogeneous* if the one-point correlation tensor is constant, while the two-point correlation tensor depends only upon the difference $\vec{y} - \vec{x}$.

Let G be a *symmetry group*, i.e., a closed subgroup of the orthogonal group $\mathrm{O}(3)$. Assume that V is an invariant subspace of the representation $g \mapsto g^{\otimes r}$ of the group G, and let $U(g)$ be the restriction of the above representation to V. A random field satisfying

$$\langle \mathsf{T}(g\vec{x}) \rangle = U(g) \langle \mathsf{T}(\vec{x}) \rangle, \qquad \langle \mathsf{T}(g\vec{x}), \mathsf{T}(g\vec{y}) \rangle = (U \otimes U)(g) \langle \mathsf{T}(\vec{x}), \mathsf{T}(\vec{y}) \rangle$$

is called *wide-sense* (G, U)*-isotropic*. In what follows we consider only wide-sense homogeneous and (G, U)-isotropic random fields and omit the words "wide-sense".

Put $\mathsf{V} = \mathsf{S}^2(\mathbb{R}^3) \subset (\mathbb{R}^3)^{\otimes 2}$, the linear space of symmetric 3×3 matrices with real entries. Consider the orthogonal representation $U(g) = \mathsf{S}^2(g)$ of the group $\mathrm{O}(3)$ as a group action: $g \cdot \vec{v} = U(g)\vec{v}$.

Mathematically, this action has 3 orbit types: orthotropic, transverse isotropic, and isotropic, see [9]. The corresponding conjugacy classes are $[H_1] = [D_2 \times Z_2^c]$, $[H_2] = [\mathrm{O}(2) \times Z_2^c]$, and $[H_3] = [\mathrm{O}(3)]$, where D_2 is the *dihedral group* of order 4 generated by the rotation about the z-axis with angle π and that about the x-axis with the same angle, and where $Z_2^c = \{I, -I\}$ with I being the identity matrix.

Physically, there are three *conductivity tensor classes* and three *stress or strain classes*: orthotropic, transverse isotropic, and isotropic.

Let $V_k = \{\,\vec{v} \in \mathsf{V}\colon\ h \cdot \vec{v} = \vec{v} \text{ for all } h \in H_k\,\}$ be the *fixed point set* of H_k. The space V_k does not depend upon the choice of a representative in the symmetry class. According to [4, Lemma 10.2], the maximal subgroup of O(3) that leaves the space V_k invariant, is the *normaliser* of H in O(3):

$$N_{\mathrm{O}(3)}(H) = \{\, g \in \mathrm{O}(3)\colon\ gHg^{-1} = H\,\}.$$

Let G be a closed subgroup of O(3) lying between H and $N_{\mathrm{O}(3)}(H)$. We will find the general form of the one-point and two-point correlation tensors of a homogeneous and (G, U)-isotropic random field, as well as the spectral expansion of the field in terms of stochastic integrals with respect to orthogonal scattered random measures.

Note that we have $N_{\mathrm{O}(3)}(D_2 \times Z_2^c) = \mathcal{O} \times Z_2^c$, $N_{\mathrm{O}(3)}(\mathrm{O}(2) \times Z_2^c) = \mathrm{O}(2) \times Z_2^c$, and $N_{\mathrm{O}(3)}(\mathrm{O}(3)) = \mathrm{O}(3)$, where $\mathcal{O}$ is the octahedral group of order 24 which fixes an octahedron. The possible values for the group G are as follows: $G_1 = D_2 \times Z_2^c$, $G_2 = D_4 \times Z_2^c$, $G_3 = \mathcal{T} \times Z_2^c$, $G_4 = \mathcal{O} \times Z_2^c$, $G_5 = \mathrm{O}(2) \times Z_2^c$, and $G_6 = \mathrm{O}(3)$, where $\mathcal{T}$ is the tetrahedral group of order 12 that fixes a tetrahedron.

10.2 The Results

First, we describe the spaces V_k and the structure of the representation U of the group G. We choose an orthonormal basis in each irreducible component of U.

The space V_1 has dimension 3 and consists of 3×3 diagonal matrices. When G is the smallest possible, $G = G_1 = D_2 \times Z_2^c$, then by definition of V_1 we have $U(g) = 3A_g$, the direct sum of three copies of the trivial irreducible representation A_g, and we choose the basis as follows:

$$T^{A_g,1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad T^{A_g,2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad T^{A_g,3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

For finite groups, we use the notation of [1, Section 4] to denote their elements and Mulliken's notation [8] or [1, Section 14] to denote their irreducible representations.

Put $G = G_2 = D_4 \times Z_2^c$. Then we have $U(g) = 2A_{1g} \oplus B_{1g}$, where A_{1g} is the trivial irreducible representation, while B_{1g} has value 1 on the elements of the set

$$G_+ = \{E, C_2, C_{21}', C_{22}', i, \sigma_h, \sigma_{v1}, \sigma_{v2}\},$$

and value -1 on the remaining elements of G. The basis is as follows:

$$T^{A_{1g},1} = T^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad T^{A_{1g},2} = T^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$T^{B_{1g},1} = T^3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Put $G = G_3 = \mathcal{T} \times Z_2^c$. Then we have $U(g) = A_g \oplus E_g$, where A_g is the trivial irreducible representation of G. To describe E_g, let the group G_3 be the union of the following three nonintersecting subsets:

$$\begin{aligned} G_0 &= \{E, C_{2x}, C_{2y}, C_{2z}, i, \sigma_x, \sigma_y, \sigma_z\}, \\ G_+ &= \{C_{31}^+, C_{32}^+, C_{33}^+, C_{34}^+, S_{61}^-, S_{62}^-, S_{63}^-, S_{64}^-\}, \\ G_- &= \{C_{31}^-, C_{32}^-, C_{33}^-, C_{34}^-, S_{61}^+, S_{62}^+, S_{63}^+, S_{64}^+\}. \end{aligned}$$

The representation E_g maps all elements of G_0 to the identity matrix, all elements of G_+ to the matrix $\frac{1}{2}\left(\begin{smallmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{smallmatrix}\right)$, and all elements of G_- to the matrix $\frac{1}{2}\left(\begin{smallmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{smallmatrix}\right)$. The basis is as follows:

$$\begin{gathered} T^1 = T^{E_g,1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad T^2 = T^{A_g} = \frac{1}{\sqrt{3}} I, \\ T^3 = T^{E_g,2} = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \end{gathered} \tag{10.2}$$

Put $G = G_4 = \mathcal{O} \times Z_2^c$. Then we have $U(g) = A_{1g} \oplus E_g$, where A_{1g} is the trivial representation of G. To describe E_g, let the group G_4 be the union of the following six nonintersecting subsets:

$$\begin{aligned} G_0 &= \{E, C_{2x}, C_{2y}, C_{2z}, i, \sigma_x, \sigma_y, \sigma_z\}, \\ G_+ &= \{C_{31}^+, C_{32}^+, C_{33}^+, C_{34}^+, S_{61}^-, S_{62}^-, S_{63}^-, S_{64}^-\}, \\ G_- &= \{C_{31}^-, C_{32}^-, C_{33}^-, C_{34}^-, S_{61}^+, S_{62}^+, S_{63}^+, S_{64}^+\}, \\ G_x &= \{C_{4x}^+, C_{4x}^-, C_{2d}', C_{2f}', S_{4x}^-, S_{4x}^+, \sigma_{d4}, \sigma_{d6}\}, \\ G_y &= \{C_{4y}^+, C_{4y}^-, C_{2c}', C_{2e}', S_{4y}^-, S_{4y}^+, \sigma_{d3}, \sigma_{d5}\}, \\ G_z &= \{C_{4z}^+, C_{4z}^-, C_{2a}', C_{2b}', S_{4z}^-, S_{4z}^+, \sigma_{d1}, \sigma_{d2}\}. \end{aligned} \tag{10.3}$$

The representation E_g maps all elements of G_0 to the identity matrix, all elements of G_+ to the matrix $\frac{1}{2}\left(\begin{smallmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{smallmatrix}\right)$, all elements of G_- to the matrix $\frac{1}{2}\left(\begin{smallmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{smallmatrix}\right)$, all elements of G_x to the matrix $\frac{1}{2}\left(\begin{smallmatrix} -1 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{smallmatrix}\right)$, all elements of G_y to the matrix $\frac{1}{2}\left(\begin{smallmatrix} -1 & \sqrt{3} \\ \sqrt{3} & 1 \end{smallmatrix}\right)$, and all elements of G_z to the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. The basis is described by Eq. (10.2).

The space V_2 has dimension 2. By definition of V_2 we have $U(g) = 2A$, the direct sum of two copies of the trivial representation A of the group $G_4 = \mathrm{O}(2) \times Z_2^c$, and we choose the basis as follows:

$$T^{A,1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad T^{A,2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Finally, the space V_3 is one-dimensional and is generated by the basis tensor $T = \frac{1}{\sqrt{3}} I$.

Second, we describe the orbit space $\hat{\mathbb{R}}^3/G_k$ for the action of each group G_k in the *wavenumber domain* $\hat{\mathbb{R}}^3$. It is stratified into a disjoint union of manifolds, say $(\hat{\mathbb{R}}^3/G_k)_m$. For the subgroups of the group $\mathrm{O}(2) \times Z_2^c$, it is convenient to use cylindrical coordinates (ρ, φ, p_3): When $G = G_1 = D_2 \times Z_2^c$, we have $\hat{\mathbb{R}}^3/D_2 \times Z_2^c = \{\, (\rho, \varphi, p_3) \colon \rho \geq 0, 0 \leq \varphi \leq \pi/2, p_3 \geq 0 \,\}$ and

$$\begin{aligned}
(\hat{\mathbb{R}}^3/D_2 \times Z_2^c)_0 &= \{\, (\rho, \varphi, p_3) \colon \rho = 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_2 \times Z_2^c)_1 &= \{\, (0, 0, p_3) \colon p_3 > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_2 \times Z_2^c)_2 &= \{\, (\rho, 0, 0) \colon \rho > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_2 \times Z_2^c)_3 &= \{\, (\rho, \pi/2, 0) \colon \rho > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_2 \times Z_2^c)_4 &= \{\, (\rho, \varphi, 0) \colon \rho > 0, 0 < \varphi < \pi/2 \,\}, \\
(\hat{\mathbb{R}}^3/D_2 \times Z_2^c)_5 &= \{\, (\rho, 0, p_3) \colon \rho > 0, p_3 > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_2 \times Z_2^c)_6 &= \{\, (\rho, \pi/2, p_3) \colon \rho > 0, p_3 > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_2 \times Z_2^c)_7 &= \{\, (\rho, \varphi, p_3) \colon \rho > 0, 0 < \varphi < \pi/2, p_3 > 0 \,\}.
\end{aligned}$$

When $G = G_2 = D_4 \times Z_2^c$, we have $\hat{\mathbb{R}}^3/D_4 \times Z_2^c = \{\, (\rho, \varphi, p_3) \colon \rho \geq 0, 0 \leq \varphi \leq \pi/4, p_3 \geq 0 \,\}$ and

$$\begin{aligned}
(\hat{\mathbb{R}}^3/D_4 \times Z_2^c)_0 &= \{\, (\rho, \varphi, p_3) \colon \rho = 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_4 \times Z_2^c)_1 &= \{\, (0, 0, p_3) \colon p_3 > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_4 \times Z_2^c)_2 &= \{\, (\rho, 0, 0) \colon \rho > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_4 \times Z_2^c)_3 &= \{\, (\rho, \pi/4, 0) \colon \rho > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_4 \times Z_2^c)_4 &= \{\, (\rho, \varphi, 0) \colon \rho > 0, 0 < \varphi < \pi/4 \,\}, \\
(\hat{\mathbb{R}}^3/D_4 \times Z_2^c)_5 &= \{\, (\rho, 0, p_3) \colon \rho > 0, p_3 > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_4 \times Z_2^c)_6 &= \{\, (\rho, \pi/4, p_3) \colon \rho > 0, p_3 > 0 \,\}, \\
(\hat{\mathbb{R}}^3/D_4 \times Z_2^c)_7 &= \{\, (\rho, \varphi, p_3) \colon \rho > 0, 0 < \varphi < \pi/4, p_3 > 0 \,\}.
\end{aligned}$$

In contrast to the previous cases, the orbit space $\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c$ is not closed. It is stratified as

$$(\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c)_0 = \{\vec{0}\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c)_1 = \{\,(0, p_3, 0)\colon p_3 > 0\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c)_2 = \{\,(p_1, p_2, p_3)\colon 0 < p_1 = p_2 = p_3\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c)_3 = \{\,(p_1, 0, p_3)\colon p_1 > 0, p_3 > 0\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c)_4 = \{\,(p_1, p_2, p_3)\colon 0 < p_1 = p_2 < p_3\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c)_5 = \{\,(p_1, p_2, p_3)\colon p_1 > 0, 0 < p_2 \le \max\{p_1, p_3\}, p_3 > 0\,\}.$$

The orbit space $\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c$ is stratified as follows:

$$(\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c)_0 = \{\vec{0}\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c)_1 = \{\,(0, 0, p_3)\colon p_3 > 0\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c)_2 = \{\,(p_1, p_2, p_3)\colon p_1 = p_2 = p_3 > 0\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c)_3 = \{\,(0, p_2, p_3)\colon 0 < p_2 = p_3\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c)_4 = \{\,(0, p_2, p_3)\colon 0 < p_2 < p_3\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c)_5 = \{\,(p_1, p_2, p_3)\colon 0 < p_1 = p_2 < p_3\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c)_6 = \{\,(p_1, p_2, p_3)\colon 0 < p_1 < p_2 = p_3\,\},$$
$$(\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c)_7 = \{\,(p_1, p_2, p_3)\colon 0 < p_1 < p_2 < p_3\,\}.$$

The orbit space $\hat{\mathbb{R}}^3/\mathrm{O}(2) \times Z_2^c$ is stratified as

$$(\hat{\mathbb{R}}^3/\mathrm{O}(2) \times Z_2^c)_0 = \{\vec{0}\},$$
$$(\hat{\mathbb{R}}^3/\mathrm{O}(2) \times Z_2^c)_1 = \{\,(p_1, 0, 0)\colon p_1 > 0\,\},$$
$$(\hat{\mathbb{R}}^3/\mathrm{O}(2) \times Z_2^c)_2 = \{\,(0, 0, p_3)\colon p_3 > 0\,\},$$
$$(\hat{\mathbb{R}}^3/\mathrm{O}(2) \times Z_2^c)_3 = \{\,(p_1, 0, p_3)\colon p_1 > 0, p_3 > 0\,\}.$$

Finally, the orbit space $\hat{\mathbb{R}}^3/\mathrm{O}(3)$ is stratified as

$$(\hat{\mathbb{R}}^3/\mathrm{O}(3))_0 = \{\vec{0}\},$$
$$(\hat{\mathbb{R}}^3/\mathrm{O}(3))_1 = \{\,(0, 0, p_3)\colon p_3 > 0\,\}.$$

Theorem 10.1 *The two-point correlation tensor of a homogeneous and* $(D_2 \times Z_2^c, 3A_g)$*-isotropic random field has the form*

$$\langle T(\vec{x}), T(\vec{y})\rangle = \int_{\hat{\mathbb{R}}^3/D_2\times Z_2^c} \cos(p_1(y_1-x_1))\cos(p_2(y_2-x_2))\cos(p_3(y_3-x_3)) \times f(\vec{p})\,\mathrm{d}\Phi(\vec{p}),$$

where $f(\vec{p})$ is a Φ-equivalence class of measurable functions acting from $\hat{\mathbb{R}}^3/D_2\times Z_2^c$ to the set of nonnegative-definite symmetric linear operators on V_1 with unit trace. The field has the form

$$T(\vec{x}) = \sum_{k=1}^{3} C_k E^{A,k} + \sum_{l=1}^{8}\sum_{k=1}^{3}\int_{\hat{\mathbb{R}}^3/D_2\times Z_2^c} v_l(\vec{p},\vec{x})\,\mathrm{d}Z^{kl}(\vec{p})T^{A,k},$$

where $C_k\in\mathbb{R}$, $v_l(\vec{p},\vec{x})$ are 8 different combinations of cosines and sines of p_1x_1, p_2x_2, and p_3x_3, and $\vec{Z}^l(\vec{p}) = (Z^{1l}(\vec{p}),\dots,Z^{3l}(\vec{p}))^\top$ are 8 centred real-valued uncorrelated random measures on $\hat{\mathbb{R}}^3/D_2\times Z_2^c$ with control measure $f(\vec{p})\,\mathrm{d}\Phi(\vec{p})$.

Put $\vec{z} = \vec{y}-\vec{x}$. Let $f^+(\vec{p})$ be a Φ-equivalence class of measurable functions acting from $\hat{\mathbb{R}}^3/D_4\times Z_2^c$ to the set of nonnegative-definite symmetric 3×3 matrices with unit trace. Let $f^-(\vec{p})$ be the same class where all non-diagonal elements but $f_{12}^+(\vec{p})$ and $f_{21}^+(\vec{p})$ are multiplied by -1. Let $f_0(\vec{p})$ be the same class where all non-diagonal elements but $f_{12}^+(\vec{p})$ and $f_{21}^+(\vec{p})$ are replaced by zeroes. Finally, let $(\hat{\mathbb{R}}^3/D_4\times Z_2^c)_{i_1,\dots,i_k}$ denote the union of the sets $(\hat{\mathbb{R}}^3/D_4\times Z_2^c)_{i_j}$ over $1\le j\le k$.

Theorem 10.2 *The two-point correlation tensor of a homogeneous and isotropic random field with $G = D_4\times Z_2^c$ and $U = 2A_{1g}\oplus B_{1g}$ has the form*

$$\begin{aligned}\langle T(\vec{x}), T(\vec{y})\rangle = {}& \frac{1}{2}\int_{(\hat{\mathbb{R}}^3/D_4\times Z_2^c)_{2,4,5,7}} \cos(p_3z_3)[\cos(p_1z_1)\cos(p_2z_2)f_+(\vec{p}) \\ & + \cos(p_1z_2)\cos(p_2z_1)f_-(\vec{p})]\,\mathrm{d}\Phi(\vec{p}) \\ & + \frac{1}{2}\int_{(\hat{\mathbb{R}}^3/D_4\times Z_2^c)_{0,1,3,6}} \cos(p_3z_3)[\cos(p_1z_1)\cos(p_2z_2) \\ & + \cos(p_1z_2)\cos(p_2z_1)]f_0(\vec{p})\,\mathrm{d}\Phi(\vec{p}).\end{aligned}$$

The field has the form

$$\begin{aligned}T(\vec{x}) = {}& \sum_{k=1}^{2} C_k E^{A_{1g},k} + \frac{1}{\sqrt{2}}\sum_{l=1}^{8}\sum_{k=1}^{3}\int_{(\hat{\mathbb{R}}^3/D_4\times Z_2^c)_{2,4,5,7}} v_l(\vec{p},\vec{x})\,\mathrm{d}Z_+^{kl}(\vec{p})T^{A,k} \\ & + \frac{1}{\sqrt{2}}\sum_{l=9}^{16}\sum_{k=1}^{3}\int_{(\hat{\mathbb{R}}^3/D_4\times Z_2^c)_{2,4,5,7}} v_l(\vec{p},\vec{x})\,\mathrm{d}Z_-^{kl}(\vec{p})T^{A,k} \\ & + \frac{1}{\sqrt{2}}\sum_{l=1}^{16}\sum_{k=1}^{3}\int_{(\hat{\mathbb{R}}^3/D_4\times Z_2^c)_{0,1,3,6}} v_l(\vec{p},\vec{x})\,\mathrm{d}Z_0^{kl}(\vec{p})T^{A,k},\end{aligned}$$

where $C_k \in \mathbb{R}$, $v_l(\vec{p},\vec{x})$, $1 \le l \le 8$ *are* 8 *different combinations of cosines and sines of* p_1x_1, p_2x_2, *and* p_3x_3, $v_l(\vec{p},\vec{x})$, $9 \le l \le 16$ *are* 8 *different combinations of cosines and sines of* p_1x_2, p_2x_1, *and* p_3x_3, $\vec{Z}^l_+(\vec{p}) = (Z^{1l}_+(\vec{p}),\dots,Z^{3l}_+(\vec{p}))^\top$, $1 \le l \le 8$ (*resp.* $\vec{Z}^l_-(\vec{p}) = (Z^{1l}_-(\vec{p}),\dots,Z^{3l}_-(\vec{p}))^\top$, $9 \le l \le 16$, *resp.* $\vec{Z}^l_0(\vec{p}) = (Z^{1l}_0(\vec{p}),\dots,Z^{3l}_0(\vec{p}))^\top$, $1 \le l \le 16$) *are* 8 *centred real-valued uncorrelated random measures on* $\hat{\mathbb{R}}^3/D_4 \times Z_2^c$ *with control measure* $f_+(\vec{p})\,\mathrm{d}\Phi(\vec{p})$ (*resp.* $f_-(\vec{p})\,\mathrm{d}\Phi(\vec{p})$, *resp.* $f_0(\vec{p})\,\mathrm{d}\Phi(\vec{p})$).

Let $f(\vec{p})$ be a Φ-equivalence class of measurable functions acting from $\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c$ to the set of nonnegative-definite symmetric 3×3 matrices with unit trace of the form

$$f(\vec{p}) = \begin{pmatrix} f_{A,2}(\vec{p}) + f_{E,2,1}(\vec{p}) & f_{E,1,1}(\vec{p}) & f_{E,2,2}(\vec{p}) \\ f_{E,1,1}(\vec{p}) & f_{A,1}(\vec{p}) & f_{E,1,2}(\vec{p}) \\ f_{E,2,2}(\vec{p}) & f_{E,1,2}(\vec{p}) & f_{A,2}(\vec{p}) - f_{E,2,1}(\vec{p}) \end{pmatrix}. \tag{10.4}$$

Let $f^0(\vec{p})$ be the same class where all non-diagonal elements are replaced by zeroes. Let $f^+(\vec{p})$ (resp. $f^-(\vec{p})$) be the same class with

$$(f^\pm_{E,i,1}(\vec{p}),\, f^\pm_{E,i,2}(\vec{p}))^\top = E_g(C^\pm_{31})(f_{E,i,1}(\vec{p}),\, f_{E,i,2}(\vec{p}))^\top$$

and $f^\pm_{A,i}(\vec{p}) = f_{A,i}(\vec{p})$. Denote

$$\begin{aligned} A^0(\vec{p},\vec{z}) &= 8\cos(p_1z_1)\cos(p_2z_2)\cos(p_3z_3), \\ A^+(\vec{p},\vec{z}) &= 8\cos(p_1z_2)\cos(p_2z_3)\cos(p_3z_1), \\ A^-(\vec{p},\vec{z}) &= 8\cos(p_1z_3)\cos(p_2z_1)\cos(p_3z_2). \end{aligned}$$

Theorem 10.3 *The two-point correlation tensor of a homogeneous and isotropic random field with* $G = \mathcal{T} \times Z_2^c$ *and* $U = A_g \oplus E_g$ *has the form*

$$\begin{aligned} \langle T(\vec{x}), T(\vec{y})\rangle = {} & \frac{1}{3}\int_{(\hat{\mathbb{R}}^3/\mathcal{T}\times Z_2^c)_{1,3-5}} \sum_{k\in\{0,+,-\}} A^k(\vec{p},\vec{z}) f^k(\vec{p})\,\mathrm{d}\Phi(\vec{p}) \\ & + \frac{1}{3}\int_{(\hat{\mathbb{R}}^3/\mathcal{T}\times Z_2^c)_{0,2}} (A^0(\vec{p}) + A^+(\vec{p}) + A^-(\vec{p}))f^0(\vec{p})\,\mathrm{d}\Phi(\vec{p}), \end{aligned}$$

The field has the form

$$\begin{aligned} T(\vec{x}) = {} & CT^2 + \frac{1}{\sqrt{3}}\sum_{l=1}^{8}\sum_{k\in\{0,+,-\}}\sum_{m=1}^{3}\int_{(\hat{\mathbb{R}}^3/\mathcal{T}\times Z_2^c)_{1,3-5}} v_{kl}(\vec{p},\vec{x})\,\mathrm{d}Z_1^{klm}(\vec{p})T^m \\ & + \frac{1}{\sqrt{3}}\sum_{l=1}^{8}\sum_{k\in\{0,+,-\}}\sum_{m=1}^{3}\int_{(\hat{\mathbb{R}}^3/\mathcal{T}\times Z_2^c)_{0,2}} v_{kl}(\vec{p},\vec{x})\,\mathrm{d}Z_2^{klm}(\vec{p})T^m, \end{aligned}$$

where $C \in \mathbb{R}$, $v_{kl}(\vec{p}, \vec{x})$, $k \in \{0, +, -\}$ *are* 8 *different combinations of cosines and sines of the terms of* $A^k(\vec{p}, \vec{z})$, *and* $\vec{Z}_1^{kl}(\vec{p}) = (Z_1^{kl1}(\vec{p}), \dots, Z_1^{kl3}(\vec{p}))^\top$ (*resp.* $\vec{Z}_2^{kl}(\vec{p}) = (Z_2^{kl1}(\vec{p}), \dots, Z_2^{kl3}(\vec{p}))^\top$) *are* 8 *centred real-valued uncorrelated random measures on* $(\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c)_{1,3-5}$ (*resp. on* $(\hat{\mathbb{R}}^3/\mathcal{T} \times Z_2^c)_{0,2}$) *with control measure* $f^k(\vec{p})\,\mathrm{d}\Phi(\vec{p})$ (*resp.* $f^0(\vec{p})\,\mathrm{d}\Phi(\vec{p})$).

Let $f^0(\vec{p})$ be a Φ-equivalence class of measurable functions acting from $\hat{\mathbb{R}}^3/\mathcal{O} \times Z_2^c$ to the convex compact set $\mathcal{C}_0$ of nonnegative-definite symmetric 3×3 matrices with unit trace of the form (10.4). Choose an element in each of the sets (10.3) and form a set $\mathcal{G}$, for example $\mathcal{G} = \{E, C_{31}^+, C_{31}^-, C_{4x}^+, C_{4y}^+, C_{4z}^+\}$. Denote by $f_h^0(\vec{p})$ the matrix $f^0(\vec{p})$, where vector $(f_{E,1,1}(\vec{p}), f_{E,1,2}(\vec{p}))^\top$ (resp. $(f_{E,2,1}(\vec{p}), f_{E,2,2}(\vec{p}))^\top$) is replaced with $E_g(h)(f_{E,1,1}(\vec{p}), f_{E,1,2}(\vec{p}))^\top$ (resp. $E_g(h)(f_{E,2,1}(\vec{p}), f_{E,2,2}(\vec{p}))^\top$), $h \in \mathcal{G}$.

Denote by $\mathcal{C}_1$ the convex compact set of all 3×3 symmetric nonnegative-definite matrices (10.4) satisfying

$$(\sqrt{3} - 2) f_{E,1,1}(\vec{p}) + f_{E,1,2}(\vec{p}) = 0,$$
$$(\sqrt{3} - 2) f_{E,2,1}(\vec{p}) + f_{E,2,2}(\vec{p}) = 0.$$

Let the matrix $f^1(\vec{p})$ takes values in $\mathcal{C}_1$. Denote by $f_h^1(\vec{p})$ the matrix $f^1(\vec{p})$, where the vector $(f_{E,1,1}(\vec{p}), f_{E,1,2}(\vec{p}))^\top$ (resp. $(f_{E,2,1}(\vec{p}), f_{E,2,2}(\vec{p}))^\top$) is replaced with $E_g(h)(f_{E,1,1}(\vec{p}), f_{E,1,2}(\vec{p}))^\top$ (resp. $E_g(h)(f_{E,2,1}(\vec{p}), f_{E,2,2}(\vec{p}))^\top$), $h \in \mathcal{G}$.

Denote by $\mathcal{C}_2$ the convex compact set of all 3×3 symmetric nonnegative-definite matrices (10.4) satisfying

$$f_{E,1,1}(\vec{p}) + f_{E,1,2}(\vec{p}) = 0,$$
$$f_{E,2,1}(\vec{p}) + f_{E,2,2}(\vec{p}) = 0.$$

Let the matrix $f^2(\vec{p})$ takes values in $\mathcal{C}_2$. Denote by $f_h^2(\vec{p})$ the matrix $f^2(\vec{p})$, where the vector $(f_{E,1,1}(\vec{p}), f_{E,1,2}(\vec{p}))^\top$ (resp. $(f_{E,2,1}(\vec{p}), f_{E,2,2}(\vec{p}))^\top$) is replaced with $E_g(h)(f_{E,1,1}(\vec{p}), f_{E,1,2}(\vec{p}))^\top$ (resp. $E_g(h)(f_{E,2,1}(\vec{p}), f_{E,2,2}(\vec{p}))^\top$), $h \in \mathcal{G}$.

Denote by $\mathcal{C}_3$ the convex compact set of all 3×3 symmetric nonnegative-definite matrices (10.4) satisfying $f_{E,i,j}(\vec{p}) = 0$. Let the matrix $f^3(\vec{p})$ takes values in $\mathcal{C}_3$.

Denote

$$A^E(\vec{p}, \vec{z}) = 8\cos(p_1 z_1)\cos(p_2 z_2)\cos(p_3 z_3),$$
$$A^{C_{31}^+}(\vec{p}, \vec{z}) = 8\cos(p_1 z_2)\cos(p_2 z_3)\cos(p_3 z_1),$$
$$A^{C_{31}^-}(\vec{p}, \vec{z}) = 8\cos(p_1 z_3)\cos(p_2 z_1)\cos(p_3 z_2),$$

$$A^{C_{4x}^+}(\vec{p}, \vec{z}) = 8\cos(p_1 z_1)\cos(p_2 z_3)\cos(p_3 z_2),$$
$$A^{C_{4y}^+}(\vec{p}, \vec{z}) = 8\cos(p_1 z_3)\cos(p_2 z_1)\cos(p_3 z_1),$$
$$A^{C_{4z}^+}(\vec{p}, \vec{z}) = 8\cos(p_1 z_2)\cos(p_2 z_1)\cos(p_3 z_3).$$

Theorem 10.4 *The two-point correlation tensor of a homogeneous and isotropic random field with $G = \mathcal{O} \times Z_2^c$ and $U = A_{1g} \oplus E_g$ has the form*

$$\begin{aligned}\langle T(\vec{x}), T(\vec{y})\rangle = \frac{1}{6}\int_{(\hat{\mathbb{R}^3}/\mathcal{O}\times Z_2^c)_{4,7}} \sum_{h\in\mathcal{G}} A^h(\vec{p},\vec{z}) f_h^0(\vec{p})\,\mathrm{d}\Phi(\vec{p}) \\ + \frac{1}{6}\int_{(\hat{\mathbb{R}^3}/\mathcal{O}\times Z_2^c)_{3,6}} \sum_{h\in\mathcal{G}} A^h(\vec{p},\vec{z}) f_h^1(\vec{p})\,\mathrm{d}\Phi(\vec{p}) \\ + \frac{1}{6}\int_{(\hat{\mathbb{R}^3}/\mathcal{O}\times Z_2^c)_{1,5}} \sum_{h\in\mathcal{G}} A^h(\vec{p},\vec{z}) f_h^2(\vec{p})\,\mathrm{d}\Phi(\vec{p}) \\ + \frac{1}{6}\int_{(\hat{\mathbb{R}^3}/\mathcal{O}\times Z_2^c)_{0,2}} \sum_{h\in\mathcal{G}} A^h(\vec{p},\vec{z}) f^3(\vec{p})\,\mathrm{d}\Phi(\vec{p}).\end{aligned}$$

The field has the form

$$\begin{aligned}T(\vec{x}) = CT^2 + \frac{1}{\sqrt{6}}\sum_{l=1}^{8}\sum_{h\in\mathcal{G}}\sum_{m=1}^{3}\int_{(\hat{\mathbb{R}^3}/\mathcal{O}\times Z_2^c)_{4,7}} v_{hl}(\vec{p},\vec{x})\,\mathrm{d}Z_{hlm}^0(\vec{p}) \\ + \frac{1}{\sqrt{6}}\sum_{l=1}^{8}\sum_{h\in\mathcal{G}}\sum_{m=1}^{3}\int_{(\hat{\mathbb{R}^3}/\mathcal{O}\times Z_2^c)_{3,6}} v_{hl}(\vec{p},\vec{x})\,\mathrm{d}Z_{hlm}^1(\vec{p}) \\ + \frac{1}{\sqrt{6}}\sum_{l=1}^{8}\sum_{h\in\mathcal{G}}\sum_{m=1}^{3}\int_{(\hat{\mathbb{R}^3}/\mathcal{O}\times Z_2^c)_{1,5}} v_{hl}(\vec{p},\vec{x})\,\mathrm{d}Z_{hlm}^2(\vec{p}) \\ + \frac{1}{\sqrt{6}}\sum_{l=1}^{8}\sum_{h\in\mathcal{G}}\sum_{m=1}^{3}\int_{(\hat{\mathbb{R}^3}/\mathcal{O}\times Z_2^c)_{0,2}} v_{hl}(\vec{p},\vec{x})\,\mathrm{d}Z_{hlm}^3(\vec{p}),\end{aligned}$$

where $\vec{Z}_{hl}^i(\vec{p}) = (Z_{hl1}^i(\vec{p}), \ldots, Z_{hl3}^i(\vec{p}))^\top$ are centred real-valued uncorrelated random measures on the sets of integration with control measures $f_h^i(\vec{p})\,\mathrm{d}\Phi(\vec{p})$, and $v_{hl}(\vec{p},\vec{x})$ are 8 different combinations of sines and cosines of the terms of $A^h(\vec{p},\vec{z})$.

Theorem 10.5 *The two-point correlation tensor of a homogeneous and* $(\mathrm{O}(2) \times Z_2^c, 2A)$*-isotropic random field has the form*

$$\begin{aligned}\langle T(\vec{x}), T(\vec{y})\rangle = 2\int_0^\infty\int_0^\infty J_0\left(\lambda\sqrt{(y_1-x_1)^2+(y_2-x_2)^2}\right)\cos(p_3(y_3-x_3)) \\ \times f(\lambda, p_3)\,\mathrm{d}\Phi(\lambda, p_3),\end{aligned}$$

where Φ is a finite Borel measure on $[0,\infty)^2$, and $f(\lambda, p_3)$ is a Φ-equivalence class of measurable functions on $[0,\infty)^2$ with values in the compact set of all nonnegative-definite linear operators in the space V_2 with unit trace. The field has the form

$$
\begin{aligned}
T(r,\varphi,z) = {} & C_1 T^{A,1} + C_2 T^{A,2} + \\
& + \sum_{m=1}^{2} \int_0^\infty \int_0^\infty J_0(\lambda r)[\cos(p_3 z)\,\mathrm{d}Z^{01m}(\lambda,\, p_3) T^{A,1} \\
& + \sin(p_3 z)\,\mathrm{d}Z^{02m}(\lambda,\, p_3) T^{A,2}] \\
& + \sqrt{2} \sum_{\ell=1}^{\infty} \sum_{m=1}^{2} \int_0^\infty \int_0^\infty J_\ell(\lambda r)[\cos(p_3 z)\cos(\ell\varphi)\,\mathrm{d}Z^{\ell 1m}(\lambda,\, p_3) T^{A,m} \\
& + \cos(p_3 z)\sin(\ell\varphi)\,\mathrm{d}Z^{\ell 2m}(\lambda,\, p_3) T^{A,m} \\
& + \sin(p_3 z)\cos(\ell\varphi)\,\mathrm{d}Z^{\ell 3m}(\lambda,\, p_3) T^{A,m} \\
& + \sin(p_3 z)\sin(\ell\varphi)\,\mathrm{d}Z^{\ell 4m}(\lambda,\, p_3) T^{A,m}],
\end{aligned}
$$

where C_1 and C_2 are arbitrary real numbers, J_ℓ are the Bessel functions, and $\vec{Z}^{\ell i} = (Z^{\ell i 1},\, Z^{\ell i 2})^\top$ are centred V_2-valued uncorrelated random measures on $[0,\infty)^2$ with control measure $f(\lambda,\, p_3)\,\mathrm{d}\Phi(\lambda,\, p_3)$.

Theorem 10.6 *The two-point correlation tensor of a homogeneous and* (O(3), *A*)*-isotropic random field has the form*

$$
\langle T(\vec{x}),\, T(\vec{y})\rangle = \int_0^\infty \frac{\sin(\lambda\|\vec{y}-\vec{x}\|)}{\lambda\|\vec{y}-\vec{x}\|}\,\mathrm{d}\Phi(\lambda).
$$

The field has the form

$$
T(\vec{x}) = \left(C + \pi\sqrt{2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} S_\ell^m(\theta,\varphi) \int_0^\infty \frac{J_{\ell+1/2}(\lambda\rho)}{\sqrt{\lambda\rho}}\,\mathrm{d}Z_\ell^m(\lambda) \right) T.
$$

10.3 A Sketch of Proofs

The general form of one- and two-point correlation tensors follow from [6, Theorem 0]. The spectral expansion of the field is obtained by using Karhunen's theorem, see [5]. The details may be found in [6, p. 200] and in the forthcoming book [7].

If one replaces the space $\mathsf{S}^2(\mathbb{R}^3)$ of the stress or strain tensors with the space $\mathsf{S}^2(\mathsf{S}^2(\mathbb{R}^3))$ of *elasticity tensors*, then there are 8 elasticity classes, see [2]. The spectral expansions of the corresponding random fields have been found by the authors in [6].

In both cases the group G is of type II, that is, it contains $-I$. If we consider the space $\mathsf{S}^2(\mathbb{R}^3) \otimes \mathbb{R}^3$ of *piezoelectricity tensors*, the situation becomes more sophisticated. There are 16 piezoelectricity classes, see [3]. Moreover, some of the groups

G are of types I or III, that is, they either are subgroups of the group SO(3) or do not contain $-I$. For such groups, Theorem 0 of [6] is no longer valid and should be replaced by another statement. The paper is under construction and will be published elsewhere. See also the forthcoming book [7].

References

1. Altmann, S.L., Herzig, P.: Point-group Theory Tables. Oxford Science Publications, Clarendon Press (1994)
2. Forte, S., Vianello, M.: Symmetry classes for elasticity tensors. J. Elast. **43**(2), 81–108 (1996). https://doi.org/10.1007/BF00042505
3. Geymonat, G., Weller, T.: Classes de symétrie des solides piézoélectriques. C. R. Math. Acad. Sci. Paris **335**(10), 847–852 (2002). https://doi.org/10.1016/S1631-073X(02)02573-6
4. Golubitsky, M., Stewart, I., Schaeffer, D.G.: Singularities and groups in bifurcation theory. Vol. II, Applied Mathematical Sciences, vol. 69. Springer, New York (1988). https://doi.org/10.1007/978-1-4612-4574-2
5. Karhunen, K.: Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys. **1947**(37), 79 (1947)
6. Malyarenko, A., Ostoja-Starzewski, M.: A random field formulation of Hooke's law in all elasticity classes. J. Elast. **127**(2), 269–302 (2017)
7. Malyarenko, A., Ostoja-Starzewski, M.: Tensor-Valued Random Fields for Continuum Physics. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge (2019)
8. Mulliken, R.S.: Electronic structures of polyatomic molecules and valence. IV. Electronic states, quantum theory of the double bond. Phys. Rev. **43**, 279–302 (1933)
9. Olive, M., Auffray, N.: Symmetry classes for even-order tensors. Math. Mech. Complex Syst. **1**(2), 177–210 (2013)

Part II
Applications of Stochastic Processes

Chapter 11
Nonlinearly Perturbed Birth-Death-Type Models

Dmitrii Silvestrov, Mikael Petersson and Ola Hössjer

Abstract Asymptotic expansions are presented for stationary and conditional quasi-stationary distributions of nonlinearly perturbed birth-death-type semi-Markov models, as well as algorithms for computing the coefficients of these expansions. Three types of applications are discussed in detail. The first is a model of population growth, where either an isolated population is perturbed by immigration, or a sink population with immigration is perturbed by internal births. The second application is epidemic spread of disease, in which a closed population is perturbed by infected individuals from outside. The third model captures the time dynamics of the genetic composition of a population with genetic drift and selection, that is perturbed by various mutation scenarios.

Keywords Semi-Markov birth-death process · Quasi-stationary distribution Nonlinear perturbation · Population dynamics model · Population genetics model Epidemic model

11.1 Introduction

Models of perturbed Markov chains and semi-Markov processes attracted attention of researchers in the mid of the 20th century, in particular the most difficult cases of perturbed processes with absorption and so-called singularly perturbed processes. An interest in these models has been stimulated by applications to control, queuing systems, information networks, and various types of biological systems. As a rule,

D. Silvestrov (✉) · O. Hössjer
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
e-mail: silvestrov@math.su.se

O. Hössjer
e-mail: ola@math.su.se

M. Petersson
Statistics Sweden, Stockholm, Sweden
e-mail: mikpe274@gmail.com

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_11

Markov-type processes with singular perturbations appear as natural tools for mathematical analysis of multi-component systems with weakly interacting components.

In this paper, we present new algorithms for construction of asymptotic expansions for stationary and conditional quasi-stationary distributions of nonlinearly perturbed semi-Markov birth-death processes with a finite phase space. We consider models that include a positive perturbation parameter that tends to zero as the unperturbed null model is approached. It is assumed that the phase space is one class of communicative states, for the embedded Markov chains of pre-limiting perturbed semi-Markov birth-death processes, whereas the limiting unperturbed model either consists of one closed class of communicative states, or of one class of communicative transient internal states that has one or both end points as absorbing states.

These new algorithms are applied to several perturbed birth-death models of biological nature. The first application is population size dynamics in a constant environment with a finite carrying capacity. It is assumed that one individual at a time is born, immigrates or dies, see, for instance, Lande, Engen and Saether [26]. In order to study the impact of immigration or births, it is possible to either view the immigration rate as a perturbation parameter of an isolated population, or the birth rate as a perturbation parameter of a sink population in which no individuals are born. The first analysis depends heavily on the ratio between the birth and death rates for the null model, whereas the second analysis involves the corresponding ratio of the immigration and death rates.

The second application is epidemic spread of a disease, reviewed, for instance, in Hethcote [14] and Nåsell [34]. Here one individual at a time gets infected or recovers, and recovered individuals become susceptible for new infections. We perturb an isolated population with no immigration, by including the possibility of occasional infected immigrants to arrive, and obtain a special case of the population dynamics model with occasional immigration.

The third application is population genetic models, treated extensively in Crow and Kimura [7] and Ewens [9]. We focus in particular on models with overlapping generations, introduced by Moran [27]. These Moran type models describe the time dynamics of the genetic composition of a population, represented as the frequency distribution of two variants of a certain gene. It is assumed that one copy of the gene is replaced for one individual at a time, and the model includes genetic drift, mutation, and various types of selection. The mutation rates between the two variants are perturbed, and the analysis depends heavily on the mutation rates and selection scheme of the unperturbed model.

The general setting of perturbed semi-Markov birth-death processes used in the paper can be motivated as follows: First, it makes it possible to consider models where inter-event times have more general non-geometric/non-exponential distributions. Second, the semi-Markov setting is a necessary element of the proposed method of sequential phase space reduction, which yields effective recurrent algorithms for computing asymptotic expansions. Third, the proposed method has a universal character. We are quite sure that it can be applied to more general models, for example, to meta-population models with several sub-populations possessing birth-death-type dynamics.

In this paper, we present asymptotic expansions of the second order and give explicit formulas for the coefficients of these expansions. The coefficients of such asymptotic expansions have a clear meaning. The first coefficients describe the asymptotic behaviour of stationary and quasi-stationary probabilities and their continuity properties with respect to small perturbations of transition characteristics of the corresponding semi-Markov birth-death processes. The second coefficients determine sensitivity of stationary and quasi-stationary probabilities with respect to small perturbations of transition characteristics.

However, it is worth to note that the proposed method can also be used for constructions of asymptotic expansions of higher orders, which also can be useful and improve accuracy of the numerical computations based on the corresponding asymptotic expansions, especially, for the models, where actual values of the perturbation parameter are not small enough to neglect the high order terms in the corresponding asymptotic expansions.

We refer here to the book by Gyllenberg and Silvestrov [13], where one can find results on asymptotic expansions for stationary and quasi-stationary distributions for perturbed semi-Markov processes, that created the background for our studies. Other recent books containing results on asymptotic expansions for perturbed Markov chains and semi-Markov processes are Korolyuk, V.S. and Korolyuk, V.V. [23], Stewart [42, 43], Konstantinov, Gu, Mehrmann and Petkov [22], Bini, Latouche and Meini [4], Koroliuk and Limnios [24], Yin and Zhang [48, 49], Avrachenkov, Filar and Howlett [3], and Silvestrov, D. and Silvestrov, S. [41]. Readers can find comprehensive bibliographies of this research area in the above books, the papers by Silvestrov, D. and Silvestrov, S. [39], Petersson [36], the doctoral dissertation of Petersson [37], and book Silvestrov, D. and Silvestrov, S. [41].

The paper includes 8 sections. In Sect. 11.2, we give examples of perturbed population dynamics, epidemic and population genetic models, which can be described in the framework of birth-death-type Markov chains and semi-Markov processes. In Sect. 11.3, we introduce a more general model of perturbed semi-Markov birth-death processes, define stationary and conditional quasi-stationary distributions for such processes and formulate basic perturbation conditions. In Sect. 11.4, we illustrate this framework for the biological models of Sect. 11.2. In Sect. 11.5, we present time-space screening procedures of phase space reduction for perturbed semi-Markov processes and recurrent algorithms for computing expectations of hitting times and stationary and conditional quasi-stationary distributions for semi-Markov birth-death processes. In Sect. 11.6, we describe algorithms for construction of the second order asymptotic expansions for stationary and conditional quasi-stationary distributions of perturbed semi-Markov birth-death processes. In Sect. 11.7, we apply the above asymptotic results to the perturbed birth-death models of biological nature defined in Sect. 11.2, and present results of related numerical studies. In Sect. 11.8, we give concluding remarks and comments.

11.2 Examples of Perturbed Birth-Death Processes

In this section, we consider a number of examples of perturbed birth-death processes that represent the time dynamics of a biological system, such as size variations of a population with a finite carrying capacity, the spread of an epidemic, or changes of the genetic composition of a population.

We let $\eta^{(\varepsilon)}(t) \in \mathbb{X} = \{0, \ldots, N\}$ denote the value of the process at time $t \geq 0$, with N a fixed (and typically large) positive integer that corresponds to the size or maximal size of the population. The perturbation parameter $\varepsilon \in (0, \varepsilon_0]$ is typically small. It either represents an immigration rate for an almost isolated population, or the mutation rate of a population in which several genetic variants segregate.

We assume that $\eta^{(\varepsilon)}(t)$ is a piecewise constant and right-continuous semi-Markov process, with discontinuities at time points

$$\zeta_n^{(\varepsilon)} = \kappa_1^{(\varepsilon)} + \cdots + \kappa_n^{(\varepsilon)}, \quad n = 0, 1, \ldots. \tag{11.1}$$

At inner points ($0 < \eta^{(\varepsilon)}(t) < N$) the process changes by one unit up or down. This either corresponds to birth or death of one individual, recovery or infection of one individual, or a change of the population's genetic decomposition. At boundary points ($\eta^{(\varepsilon)}(t) \in \{0, N\}$), any jump out of the state space is projected back to $\mathbb{X}$, so that, for instance, a "jump" from 0 ends at 0 or 1.

The time $\kappa_n^{(\varepsilon)}$ between the n:th and $(n+1)$:th jumps of $\eta^{(\varepsilon)}(t)$ will be referred to as the n:th transition time. Its distribution function

$$F_i^{(\varepsilon)}(t) = \mathsf{P}\{\kappa_n^{(\varepsilon)} \leq t / \eta^{(\varepsilon)}(\zeta_{n-1}^{(\varepsilon)}) = i\} \tag{11.2}$$

only depends on the state $i \in \mathbb{X}$ from which a jump occurs.

In this section, we consider two examples of transition time distributions (11.2). The first one is geometric,

$$F_i^{(\varepsilon)} \sim \mathrm{Ge}\,[\lambda_i(\varepsilon)] \Longrightarrow F_i^{(\varepsilon)}(t) = 1 - [1 - \lambda_i(\varepsilon)]^{[t]}, \tag{11.3}$$

where $[t]$ is the integer part of t and $0 < \lambda_i(\varepsilon) \leq 1$ represents the probability that a jump occurs in one time step. The second example corresponds to a continuous time Markov process, with an exponential transition time distribution

$$F_i^{(\varepsilon)} \sim \mathrm{Exp}\,[\lambda_i(\varepsilon)] \Longrightarrow F_i^{(\varepsilon)}(t) = 1 - e^{-\lambda_i(\varepsilon)t}, \tag{11.4}$$

with $0 < \lambda_i(\varepsilon) < \infty$ the rate at which a jump occurs. It is convenient to decompose

$$\lambda_i(\varepsilon) = \lambda_{i,-}(\varepsilon) + \lambda_{i,+}(\varepsilon) \tag{11.5}$$

as a sum of two terms, where $\lambda_{i,-}(\varepsilon)$ represents the probability of death in one time step in (11.3), or the rate at which a death occurs in (11.4) ($i \to i - 1$ when $i > 0$,

$0 \to 0$ when $i = 0$). Similarly, $\lambda_{i,+}(\varepsilon)$ is the probability or rate of a birth event ($i \to i+1$ when $i < N$, $N \to N$ when $i = N$). For both models (11.3) or (11.4), $\eta_n^{(\varepsilon)} = \eta^{(\varepsilon)}(\zeta_n^{(\varepsilon)}), \quad n = 0, 1, 2, \ldots$ is an embedded discrete time Markov chain, with transition probabilities

$$p_{i,+}(\varepsilon) = 1 - p_{i,-}(\varepsilon) = \frac{\lambda_{i,+}(\varepsilon)}{\lambda_i(\varepsilon)} \tag{11.6}$$

of jumping upwards or downwards.

It is assumed that $\mathbb{X}$ is one single class of communicative states for each $\varepsilon > 0$. The behaviour of the limiting $\varepsilon = 0$ model will satisfy one of the following three conditions:

$$\begin{aligned}
&\mathbf{H}_1 : \text{The } \varepsilon = 0 \text{ model has one class } \mathbb{X} \text{ of communicative states,} \\
&\mathbf{H}_2 : \text{The } \varepsilon = 0 \text{ model has one absorbing state 0 and one class} \\
&\qquad {}_0\mathbb{X} = \mathbb{X} \setminus \{0\} \text{ of communicative transient states,} \\
&\mathbf{H}_3 : \text{The } \varepsilon = 0 \text{ model has two absorbing states 0 and } N, \text{ and one} \\
&\qquad \text{class } {}_{0,N}\mathbb{X} = \mathbb{X} \setminus \{0, N\} \text{ of communicative transient states.}
\end{aligned} \tag{11.7}$$

These three perturbation scenarios can be rephrased in terms of the birth and death rates (11.5) as follows:

$$\begin{aligned}
&\mathbf{H}_1 : \lambda_{0,+}(0) > 0, \lambda_{N,-}(0) > 0, \\
&\mathbf{H}_2 : \lambda_{0,+}(0) = 0, \lambda_{N,-}(0) > 0, \\
&\mathbf{H}_3 : \lambda_{0,+}(0) = 0, \lambda_{N,-}(0) = 0.
\end{aligned} \tag{11.8}$$

This will be utilised in Sects. 11.2.1–11.2.3 in order to characterise the various perturbed models that we propose.

11.2.1 Perturbed Population Dynamics Models

Let N denote the maximal size of a population, and let $\eta^{(\varepsilon)}(t)$ be its size at time t. In order to model the dynamics of the population, we introduce births, deaths, and immigration from outside, according to a parametric model with

$$\lambda_{i,+}(\varepsilon) = \lambda i \left[1 - \alpha_1 \left(\frac{i}{N}\right)^{\theta_1}\right] + \nu \left[1 - \left(\frac{i}{N}\right)^{\theta_2}\right] \tag{11.9}$$

and

$$\lambda_{i,-}(\varepsilon) = \mu i \left[1 + \alpha_2 \left(\frac{i}{N}\right)^{\theta_3}\right]. \tag{11.10}$$

For a small population ($i \ll N$), we interpret the three parameters $\lambda > 0$, $\mu > 0$ and $\nu > 0$ as a birth rate per individual, a death rate per individual, and an immigration rate, whereas α_k, θ_k are density regulation parameters that model decreased birth/immigration and increased death for a population close to its maximal size. They satisfy $\theta_k > 0$, $\alpha_1 \leq 1$, and $\alpha_1, \alpha_2 \geq 0$, where the last inequality is strict for at least one of α_1 and α_2. A more general model would allow birth, death, and immigration rates to vary non-parametrically with i.

The expected growth rate of the population, when $0 < i < N$, is

$$\begin{aligned}
&\mathsf{E}\left[\eta^{(\varepsilon)}(t+\Delta t) - \eta^{(\varepsilon)}(t) | \eta^{(\varepsilon)}(t) = i\right] = \Delta t \left[\lambda_{i,+}(\varepsilon) - \lambda_{i,-}(\varepsilon)\right] \\
&\quad = \Delta t \left\{\lambda i \left[1 - \alpha_1 \left(\tfrac{i}{N}\right)^{\theta_1}\right] + \nu \left[1 - \left(\tfrac{i}{N}\right)^{\theta_2}\right] - \mu i \left[1 + \alpha_2 \left(\tfrac{i}{N}\right)^{\theta_3}\right]\right\},
\end{aligned}$$

where $\Delta t = 1$ in discrete time (11.3), and $\Delta t > 0$ is infinitesimal in continuous time (11.4). When $\theta_1 = \theta_2 = \theta_3 = \theta$, this expression simplifies to

$$\begin{aligned}
&\mathsf{E}\left[\eta^{(\varepsilon)}(t+\Delta t) - \eta^{(\varepsilon)}(t) | \eta^{(\varepsilon)}(t) = i\right] \\
&\quad = \Delta t \left\{\lambda i \left[1 - \alpha_1 \left(\tfrac{i}{N}\right)^{\theta}\right] - \mu i \left[1 + \alpha_2 \left(\tfrac{i}{N}\right)^{\theta}\right]\right\} \\
&\quad + \nu \left[1 - \left(\tfrac{i}{N}\right)^{\theta}\right] \\
&\quad = \Delta t \cdot \begin{cases}
(\lambda - \mu) i \left[1 - \frac{\alpha_1 \lambda + \alpha_2 \mu}{\lambda - \mu} \left(\frac{i}{N}\right)^{\theta}\right] & \\
+ \nu \left[1 - \left(\frac{i}{N}\right)^{\theta}\right], & \text{if } \lambda \neq \mu \\
-\mu i (\alpha_1 + \alpha_2) \left(\frac{i}{N}\right)^{\theta} + \nu \left[1 - \left(\frac{i}{N}\right)^{\theta}\right], & \text{if } \lambda = \mu.
\end{cases}
\end{aligned} \tag{11.11}$$

We shall consider two perturbation scenarios. The first one has

$$\mathbf{H}_2 : \quad \nu = \nu(\varepsilon) = \varepsilon, \tag{11.12}$$

whereas all other parameters are kept fixed, not depending on ε. It is also possible to consider more general nonlinear functions $\nu(\varepsilon)$, but this will hardly add more insight to how immigration affects population dynamics.

The unperturbed $\varepsilon = 0$ model corresponds to an isolated population that only increases through birth events. For small ε, we can think of a population that resides on an island and faces subsequent extinction and recolonisation events. After the population temporarily dies out, the island occasionally receives new immigrants at rate or probability ε. We shall find in Sect. 11.4.1 that for small migration rates ε, the properties of the model are highly dependent on whether the basic reproduction number

$$R_0 = \frac{\lambda}{\mu} \tag{11.13}$$

exceeds 1 or not.

A second perturbation scenario has a birth rate

$$\mathbf{H}_1: \quad \lambda = \lambda(\varepsilon) = \varepsilon \tag{11.14}$$

that equals ε, whereas all other parameters are kept fixed, not depending on ε. Again, more general nonlinear functions $\lambda(\varepsilon)$ can be studied, but for simplicity assume that (11.14) holds. The unperturbed $\varepsilon = 0$ model corresponds to a sink population that only increases through immigration, and its properties depend heavily on ν/μ.

11.2.2 Perturbed Epidemic Models

In order to model an epidemic in a population of size N, we let $\eta^{(\varepsilon)}(t)$ refer to the number of infected individuals at time t, whereas the remaining $N - \eta^{(\varepsilon)}(t)$ are susceptible. We assume that

$$\lambda_{i,+}(\varepsilon) = \lambda i \left(1 - \frac{i}{N}\right) + \nu(N - i), \tag{11.15}$$

and

$$\lambda_{i,-}(\varepsilon) = \mu i, \tag{11.16}$$

where the first parameter $\lambda(N-1)/N \approx \langle$ is the total contact rate between each individual and the other members of the population. The first term on the right hand side of (11.15) may be written as the product of the force of infection $\lambda i/N$ caused by i infected individuals, and the number of susceptibles $N - i$. The second parameter of the model, ν, is the contact rate between each individual and the group of infected ones outside of the population. The third parameter μ is the recovery rate per individual. It may also include a combined death and birth of an infected and susceptible individual. The model in (11.15)–(11.16) is an SIS-epidemic, since infected individuals become susceptible after recovery. It is essentially a special case of (11.9)–(11.10), with $\theta_1 = \theta_2 = \theta_3 = 1, \alpha_1 = 1$ and $\alpha_2 = 0$, although immigration is parameterised differently in (11.9) and (11.15).

Assume that the external contact rate

$$\mathbf{H}_2: \quad \nu = \nu(\varepsilon) = \varepsilon \tag{11.17}$$

equals the perturbation parameter, whereas all other parameters are kept fixed, not depending on ε. The unperturbed $\varepsilon = 0$ model refers to an isolated population without external contagion. The epidemic will then, sooner or later, die out and reach the only absorbing state 0.

Weiss and Dishon [46] first formulated the SIS-model as a continuous time birth-death Markov process (11.4) without immigration ($\varepsilon = 0$). It has since then been extended in a number of directions, see, for instance, Cavender [5], Kryscio and

Lefévre [25], Jacquez and O'Neill [18], Jacquez and Simon [19], Nåsell [29, 30] and Allen and Burgin [2]. The quasi-stationary distribution of $\eta^{(\varepsilon)}(t)$ is studied in several of these papers. In this work, we generalise previously studied models of epidemic spread by treating discrete and continuous time in a unified manner through semi-Markov processes.

The expected growth rate of the null model $\varepsilon = 0$ satisfies

$$\mathsf{E}\left[\eta^{(0)}(t+\Delta t) - \eta^{(0)}(t) | \eta^{(0)}(t) = i\right] = \Delta t \cdot ri\left(1 - \frac{i}{K(0)}\right), \tag{11.18}$$

if $0 < i < N$, when the basic reproduction ratio $R_0 = \lambda/\mu$ exceeds 1. This implies that the expected number of infected individuals follows Verhulst's logistic growth model (Verhulst [45]), with intrinsic growth rate $r = \mu(R_0 - 1)$, and a carrying capacity $K(0) = N(1 - R_0^{-1})$ of the environment.

11.2.3 Perturbed Models of Population Genetics

Let N be a positive even integer, and consider a one-sex population with $N/2$ individuals, each one of which carries two copies of a certain gene. This gene exists in two variants (or alleles); A_1 and A_2. Let $\eta^{(\varepsilon)}(t)$ be the number of gene copies with allele A_1 at time t. Consequently, the remaining $N - \eta^{(\varepsilon)}(t)$ gene copies have the other allele A_2 at time t. At each moment $\zeta_n^{(\varepsilon)}$ of jump in (11.1), a new gene copy replaces an existing one, so that

$$\eta^{(\varepsilon)}(\zeta_n^{(\varepsilon)}) = \begin{cases} \eta^{(\varepsilon)}(\zeta_n^{(\varepsilon)}-) + 1, & \text{if } A_1 \text{replaces } A_2, \\ \eta^{(\varepsilon)}(\zeta_n^{(\varepsilon)}-), & \text{if } A_k \text{replaces } A_k, \\ \eta^{(\varepsilon)}(\zeta_n^{(\varepsilon)}-) - 1, & \text{if } A_2 \text{replaces } A_1. \end{cases} \tag{11.19}$$

In discrete time (11.3), we define $\lambda_{ij}(\varepsilon)$ as the probability that the number of A_1 alleles changes from i to j when a gene copy is replaced, at each time step. In continuous time (11.4), we let $\lambda_{ij}(\varepsilon)$ be the rate at which the number of A_1 alleles changes from i to j when a gene copy replacement occurs. Let x^{**} refer to the probability that the new gene copy has variant A_1 when the fraction of A_1-alleles before replacement is $x = i/N$. We further assume that the removed gene copy is chosen randomly among all N gene copies, with equal probabilities $1/N$, so that

$$\lambda_{ij}(\varepsilon) = \begin{cases} x^{**}(1-x), & j = i+1, \\ (1-x^{**})x, & j = i-1, \\ 1 - x^{**}(1-x) - (1-x^{**})x, & j = i. \end{cases} \tag{11.20}$$

Notice that in order to make $\eta^{(\varepsilon)}(t)$ a semi-Markov process of birth-death type that satisfies (11.6), we do not regard instances when the new gene copy replaces a gene copy with the same allele as a moment of jump, if the current number i of A_1 alleles

satisfies $0 < i < N$. That is, the second line on the right hand side of (11.19) is only possible in a homogeneous population where all gene copies have the same allele A_1 or A_2, and therefore $\lambda_{ii}(\varepsilon)$ is not included in the probability or rate $\lambda_i(\varepsilon)$ to leave state i in (11.5), when $0 < i < N$.

The choice of x^{**} will determine the properties of the model. The new gene copy is formed in two steps. In the first step, a pair of genes is drawn randomly with replacement, so that its genotype is A_1A_1, A_1A_2 and A_2A_2 with probabilities x^2, $2x(1-x)$ and $(1-x)^2$ respectively. Since the gene pair is drawn with replacement, this corresponds to a probability $2/N$ that the two genes originate from the same individual (self fertilisation). A gene pair survives with probabilities proportional to $1+s_1$, 1 and $1+s_2$ for these three genotypes, where $1+s_1 \geq 0$ and $1+s_2 \geq 0$ determine the fitnesses of genotypes A_1A_1 and A_2A_2 relative to that of genotype A_1A_2. This is repeated until a surviving gene pair appears, from which a gene copy is picked randomly. Consequently, the probability is

$$x^* = \frac{1 \cdot (1+s_1)x^2 + \frac{1}{2} \cdot 2x(1-x)}{(1+s_1)x^2 + 2x(1-x) + (1+s_2)(1-x)^2} \tag{11.21}$$

that the chosen allele is A_1. In the second step, before the newly formed gene copy is put into the population, an A_1 allele mutates with probability $u_1 = \mathsf{P}(A_1 \to A_2)$, and an A_2 allele with probability $u_2 = \mathsf{P}(A_2 \to A_1)$. This implies that

$$x^{**} = (1-u_1)x^* + u_2(1-x^*). \tag{11.22}$$

By inserting (11.22) into (11.20), and (11.20) into (11.6) we get a semi-Markov process of Moran type that describes the time dynamics of two alleles in a one-sex population in the presence of selection and mutation. A special case of it was originally introduced by Moran [27], and some of its properties can be found, for instance, in Karlin and McGregor [20] and Durrett [8]. The model incorporates a number of different selection scenarios. A selectively neutral model corresponds to all three genotypes having the same fitness ($s_1 = s_2 = 0$), for directional selection, one of the two alleles is more fit than the other ($s_1 < 0 < s_2$ or $s_1 > 0 > s_2$), an underdominant model has a heterozygous genotype A_1A_2 with smaller fitness than the two homozygous genotypes A_1A_1 and A_2A_2 ($s_1, s_2 > 0$), whereas overdominance or balancing selection means that the heterozygous genotype is the one with highest fitness ($s_1, s_2 < 0$).

In continuous time (11.4), the expected value of the Moran model satisfies a differential equation

$$\begin{aligned}
&\mathsf{E}\left[\eta^{(\varepsilon)}(t+\Delta t)-\eta^{(\varepsilon)}(t)|\eta^{(\varepsilon)}(t)=Nx\right]=\Delta t\left[\lambda_{i,+}(\varepsilon)-\lambda_{i,-}(\varepsilon)\right]\\
&\quad=\Delta t\left[x^{**}(1-x)-x(1-x^{**})\right]\\
&\quad=\Delta t\,(x^{**}-x)\\
&\quad=\Delta t\left[(1-u_1-u_2)x^{*}+u_2-x\right]\\
&\quad=\Delta t\left[(1-u_1-u_2)\tfrac{x+s_1x^2}{1+s_1x^2+s_2(1-x)^2}+u_2-x\right]\\
&\quad=:\Delta t\left[N^{-1}m(x)+o(N^{-1})\right],
\end{aligned}\tag{11.23}$$

whenever $0<x<1$, with $\Delta t>0$ infinitesimal. The discrete time Moran model (11.3) also satisfies (11.23), interpreted as a difference equation, with $\Delta t=1$. In the last step of (11.23), we assumed that all mutation and selection parameters are inversely proportional to population size;

$$\begin{aligned}
u_1&=U_1/N,\\
u_2&=U_2/N,\\
s_1&=S_1/N,\\
s_2&=S_2/N,
\end{aligned}\tag{11.24}$$

and introduced an infinitesimal drift function

$$m(x)=U_2(1-x)-U_1x+[(S_1+S_2)x-S_2]\,x(1-x).$$

The corresponding infinitesimal variance function $v(x)=2x(1-x)$ follows similarly from (11.24), according to

$$\begin{aligned}
&\mathsf{V}\left[\eta^{(\varepsilon)}(t+\Delta t)|\eta^{(\varepsilon)}(t)=Nx\right]=\Delta t\left[\lambda_{i,+}(\varepsilon)+\lambda_{i,-}(\varepsilon)+O(N^{-1})\right]\\
&\quad=\Delta t\left[x^{**}(1-x)+x(1-x^{**})+O(N^{-1})\right]\\
&\quad=\Delta t\left[2x(1-x)+O(u_1+u_2+|s_1|+|s_2|)+O(N^{-1})\right]\\
&\quad=:\Delta t\left[v(x)+O(N^{-1})\right].
\end{aligned}\tag{11.25}$$

Assume that N is fixed, whereas the perturbation parameter ε varies. We let the two selection parameters s_1 and s_2, and hence also the rescaled selection parameters S_1 and S_2, be independent of ε, whereas the rescaled mutation parameters satisfy

$$\begin{aligned}
U_1&=U_1(\varepsilon)=C_1+D_1\varepsilon,\\
U_2&=U_2(\varepsilon)=C_2+D_2\varepsilon,
\end{aligned}\tag{11.26}$$

for some non-negative constants C_1, D_1, C_2, D_2, where at least one of D_1 and D_2 is strictly positive. It follows from (11.8) and (11.20) that the values of $0\le C_1, C_2<1$ will determine the properties of the unperturbed $\varepsilon=0$ model, according to the three distinct scenarios

$$\begin{aligned}
&\mathbf{H}_1: C_1>0, C_2>0,\\
&\mathbf{H}_2: C_1>0, C_2=0,\\
&\mathbf{H}_3: C_1=0, C_2=0.
\end{aligned}\tag{11.27}$$

The null model $\varepsilon = 0$ incorporates two-way mutations $A_1 \to A_2$ and $A_2 \to A_1$ for Perturbation scenario $\mathbf{H}_1$, with no absorbing state, it has one-way mutations $A_1 \to A_2$ for Perturbation scenario $\mathbf{H}_2$, with $i = 0$ as absorbing state, and no mutations for Perturbation scenario $\mathbf{H}_3$, with $i = 0$ and $i = N$ as the two absorbing states.

11.3 Nonlinearly Perturbed Semi-Markov Birth-Death Processes

In this section, we will generalise the framework of Sect. 11.2 and introduce a model of perturbed semi-Markov birth-death processes, define stationary and conditional quasi-stationary distributions for such processes and formulate basic perturbation conditions.

11.3.1 Perturbed Semi-Markov Birth-Death Processes

Let $(\eta_n^{(\varepsilon)}, \kappa_n^{(\varepsilon)})$, $n = 0, 1, \ldots$ be, for every value of a perturbation parameter $\varepsilon \in (0, \varepsilon_0]$, where $0 < \varepsilon_0 \le 1$, a Markov renewal process, i.e., a homogeneous Markov chain with the phase space $\mathbb{X} \times [0, \infty)$, where $\mathbb{X} = \{0, 1, \ldots, N\}$, an initial distribution $\bar{p}^{(\varepsilon)} = \langle p_i^{(\varepsilon)} = \mathsf{P}\{\eta_0^{(\varepsilon)} = i, \kappa_0^{(\varepsilon)} = 0\} = \mathsf{P}\{\eta_0^{(\varepsilon)} = i\}, i \in \mathbb{X}\rangle$ and transition probabilities, defined for $(i, s), (j, t) \in \mathbb{X} \times [0, \infty)$,

$$Q_{ij}^{(\varepsilon)}(t) = \mathsf{P}\{\eta_1^{(\varepsilon)} = j, \kappa_1^{(\varepsilon)} \le t / \eta_0^{(\varepsilon)} = i, \kappa_0^{(\varepsilon)} = s\}$$

$$= \begin{cases} F_{0,\pm}^{(\varepsilon)}(t) p_{0,\pm}(\varepsilon) & \text{if } j = 0 + \frac{1 \pm 1}{2}, \quad \text{for } i = 0, \\ F_{i,\pm}^{(\varepsilon)}(t) p_{i,\pm}(\varepsilon) & \text{if } j = i \pm 1, \qquad \text{for } 0 < i < N, \\ F_{N,\pm}^{(\varepsilon)}(t) p_{N,\pm}(\varepsilon) & \text{if } j = N - \frac{1 \mp 1}{2}, \text{ for } i = N, \\ 0 & \text{otherwise,} \end{cases} \tag{11.28}$$

where: (a) $F_{i,\pm}^{(\varepsilon)}(t)$, $i \in \mathbb{X}$ are distribution functions concentrated on $[0, \infty)$, for every $\varepsilon \in (0, \varepsilon_0]$; (b) $p_{i,\pm}(\varepsilon) \ge 0$, $p_{i,-}(\varepsilon) + p_{i,+}(\varepsilon) = 1$, $i \in \mathbb{X}$, for every $\varepsilon \in (0, \varepsilon_0]$.

In this case, the random sequence $\eta_n^{(\varepsilon)}$ is also a homogeneous (embedded) Markov chain with the phase space $\mathbb{X}$ and the transition probabilities, defined for $i, j \in \mathbb{X}$,

$$p_{ij}(\varepsilon) = \mathsf{P}\{\eta_1^{(\varepsilon)} = j / \eta_0^{(\varepsilon)} = i\} = Q_{ij}^{(\varepsilon)}(\infty)$$

$$= \begin{cases} p_{0,\pm}(\varepsilon) & \text{if } j = 0 + \frac{1 \pm 1}{2}, \quad \text{for } i = 0, \\ p_{i,\pm}(\varepsilon) & \text{if } j = i \pm 1, \qquad \text{for } 0 < i < N, \\ p_{N,\pm}(\varepsilon) & \text{if } j = N - \frac{1 \mp 1}{2}, \text{ for } i = N, \\ 0 & \text{otherwise.} \end{cases} \tag{11.29}$$

We assume that the following condition holds:

A: $p_{i,\pm}(\varepsilon) > 0,\ i \in \mathbb{X}$, for every $\varepsilon \in (0, \varepsilon_0]$.

Condition **A** obviously implies that the phase space $\mathbb{X}$ is a communicative class of states for the embedded Markov chain $\eta_n^{(\varepsilon)}$, for every $\varepsilon \in (0, \varepsilon_0]$.

We exclude instant transitions and assume that the following condition holds:

B: $F_{i,\pm}^{(\varepsilon)}(0) = 0,\ i \in \mathbb{X}$, for every $\varepsilon \in (0, \varepsilon_0]$.

Let us now introduce a semi-Markov process,

$$\eta^{(\varepsilon)}(t) = \eta^{(\varepsilon)}_{\nu^{(\varepsilon)}(t)},\ t \geq 0, \tag{11.30}$$

where $\nu^{(\varepsilon)}(t) = \max(n \geq 0 : \zeta_n^{(\varepsilon)} \leq t)$ is the number of jumps in the time interval $[0, t]$, for $t \geq 0$, and $\zeta_n^{(\varepsilon)}$ are sequential moments of jumps for the semi-Markov process $\eta^{(\varepsilon)}(t)$. This process has the phase space $\mathbb{X}$, the initial distribution $\bar{p} = \langle p_i = \mathsf{P}\{\eta^{(\varepsilon)}(0) = i\}, i \in \mathbb{X}\rangle$ and transition probabilities $Q_{ij}^{(\varepsilon)}(t), t \geq 0, i, j \in \mathbb{X}$.

Due to the specific assumptions imposed on the transition probabilities $p_{ij}^{(\varepsilon)}, i, j \in \mathbb{X}$ in relation (11.29), one can refer to $\eta^{(\varepsilon)}(t)$ as a semi-Markov birth-death process.

If $F_{i,\pm}^{(\varepsilon)}(t) = \mathrm{I}(t \geq 1), t \geq 0, i, j \in \mathbb{X}$, then $\eta^{(\varepsilon)}(t) = \eta^{(\varepsilon)}_{[t]}, t \geq 0$ is a discrete time homogeneous Markov birth-death chain embedded in continuous time.

If $F_{ij}^{(\varepsilon)}(t) = (1 - e^{-\lambda_i(\varepsilon)t}), t \geq 0, i, j \in \mathbb{X}$ (here, $0 < \lambda_i(\varepsilon) < \infty, i \in \mathbb{X}$), then $\eta^{(\varepsilon)}(t), t \geq 0$ is a continuous time homogeneous Markov birth-death process.

Let us define expectations of transition times, for $i, j \in \mathbb{X}$,

$$e_{ij}(\varepsilon) = \mathsf{E}_i\{\kappa_1^{(\varepsilon)} I(\eta_1^{(\varepsilon)} = j)\} = \int_0^\infty t\, Q_{ij}^{(\varepsilon)}(dt) \tag{11.31}$$

$$= \begin{cases} e_{0,\pm}(\varepsilon) & \text{if } j = 0 + \frac{1\pm1}{2}, \quad \text{for } i = 0, \\ e_{i,\pm}(\varepsilon) & \text{if } j = i \pm 1, \quad\ \text{for } 0 < i < N, \\ e_{N,\pm}(\varepsilon) & \text{if } j = N - \frac{1\mp1}{2}, \text{ for } i = N, \\ 0 & \text{otherwise,} \end{cases} \tag{11.32}$$

and

$$e_i(\varepsilon) = \mathsf{E}_i\kappa_1^{(\varepsilon)} = e_{i,-}(\varepsilon) + e_{i,+}(\varepsilon). \tag{11.33}$$

Here and henceforth, the notations P_i and E_i are used for conditional probabilities and expectations under the condition $\eta^{(\varepsilon)}(0) = i$.

We also assume that the following condition holds:

C: $e_{i,\pm}(\varepsilon) < \infty,\ i, j \in \mathbb{X}$, for $\varepsilon \in (0, \varepsilon_0]$.

It is useful to note that conditions **B** and **C** imply that all expectations $e_i(\varepsilon) \in (0, \infty), i \in \mathbb{X}$.

In the case of discrete time Markov birth-death chain, $e_i(\varepsilon) = 1, i \in \mathbb{X}$, whereas in the case of continuous time Markov birth-death process, $e_i(\varepsilon) = \lambda_i^{-1}(\varepsilon), i \in \mathbb{X}$.

Conditions **A**–**C** imply that the semi-Markov birth-death process $\eta^{(\varepsilon)}(t)$ is, for every $\varepsilon \in (0, \varepsilon_0]$, ergodic in the sense that the following asymptotic relation holds,

$$\mu_i^{(\varepsilon)}(t) = \frac{1}{t}\int_0^t I(\eta^{(\varepsilon)}(s) = i)ds \xrightarrow{a.s.} \pi_i(\varepsilon) \text{ as } t \to \infty,\ i \in \mathbb{X}. \tag{11.34}$$

The ergodic relation (11.34) holds for any initial distribution $\bar{p}^{(\varepsilon)}$ and the stationary probabilities $\pi_i(\varepsilon)$, $i \in \mathbb{X}$ do not depend on the initial distribution. Moreover, $\pi_i(\varepsilon) > 0$, $i \in \mathbb{X}$ and these probabilities are the unique solution of the following system of linear equations,

$$\pi_i(\varepsilon)e_i^{-1}(\varepsilon) = \sum_{j\in\mathbb{X}} \pi_j(\varepsilon)e_j^{-1}(\varepsilon)p_{ji}(\varepsilon),\ i \in \mathbb{X},\ \sum_{i\in\mathbb{X}} \pi_i(\varepsilon) = 1. \tag{11.35}$$

11.3.2 Perturbation Conditions for Semi-Markov Birth-Death Processes

Let us assume that that the following perturbation conditions hold:

D: $p_{i,\pm}(\varepsilon) = \sum_{l=0}^{1+l_{i,\pm}} a_{i,\pm}[l]\varepsilon^l + o_{i,\pm}(\varepsilon^{1+l_{i,\pm}})$, $\varepsilon \in (0, \varepsilon_0]$, for $i \in \mathbb{X}$, where: **(a)** $|a_{i,\pm}[l]| < \infty$, for $0 \le l \le 1 + l_{i,\pm}$, $i \in \mathbb{X}$; **(b)** $l_{i,\pm} = 0$, $a_{i,\pm}[0] > 0$, for $0 < i < N$; **(c)** $l_{0,\pm} = 0$, $a_{0,\pm}[0] > 0$ or $l_{0,+} = 1$, $a_{0,+}[0] = 0$, $a_{0,+}[1] > 0$, $l_{0,-} = 0$, $a_{0,-}[0] > 0$; **(d)** $l_{N,\pm} = 0$, $a_{N,\pm}[0] \ge 0$ or $l_{N,+} = 0$, $a_{N,+}[0] > 0$, $l_{N,-} = 1$, $a_{N,-}[0] = 0$, $a_{N,-}[1] > 0$; **(e)** $o_{i,\pm}(\varepsilon^{1+l_{i,\pm}})/\varepsilon^{1+l_{i,\pm}} \to 0$ as $\varepsilon \to 0$, for $i \in \mathbb{X}$.

and

E: $e_{i,\pm}(\varepsilon) = \sum_{l=0}^{1+l_{i,\pm}} b_{i,\pm}[l]\varepsilon^l + \dot{o}_{i,\pm}(\varepsilon^{1+l_{i,\pm}})$, $\varepsilon \in (0, \varepsilon_0]$, for $i \in \mathbb{X}$, where: **(a)** $|b_{i,\pm}[l]| < \infty$, for $0 \le l \le L + l_{i,\pm}$, $i \in \mathbb{X}$; **(b)** $l_{i,\pm} = 0$, $b_{i,\pm}[0] > 0$, for $0 < i < N$; **(c)** $l_{0,\pm} = 0$, $b_{0,\pm}[0] > 0$ or $l_{0,+} = 1$, $b_{0,+}[0] = 0$, $b_{0,+}[1] > 0$, $l_{0,-} = 0$, $b_{0,-}[0] > 0$; **(d)** $l_{N,\pm} = 0$, $b_{N,\pm}[0] > 0$ or $l_{N,+} = 0$, $b_{N,+}[0] > 0$, $l_{N,-} = 1$, $b_{N,-}[0] = 0$, $b_{N,-}[1] > 0$; **(e)** $\dot{o}_{i,\pm}(\varepsilon^{1+l_{i,\pm}})/\varepsilon^{1+l_{i,\pm}} \to 0$ as $\varepsilon \to 0$, for $i \in \mathbb{X}$.

It is useful to explain the role played by the parameters $l_{i,\pm}$ in conditions **D** and **E**. These parameters equalise the so-called length of asymptotic expansions penetrating these conditions.

The length of an asymptotic expansion is defined as the number of coefficients for powers of ε in this expansion, beginning from the first non-zero coefficient and up to the coefficient for the largest power of ε in this expansion.

The asymptotic expansions penetrating conditions **D** and **E** can be rewritten in the following form, $p_{i,\pm}(\varepsilon) = \sum_{l=l_{i,\pm}}^{1+l_{i,\pm}} a_{i,\pm}[l]\varepsilon^l + o_{i,\pm}(\varepsilon^{1+l_{i,\pm}})$, $\varepsilon \in (0, \varepsilon_0]$ and $e_{i,\pm}(\varepsilon) = \sum_{l=l_{i,\pm}}^{1+l_{i,\pm}} b_{i,\pm}[l]\varepsilon^l + \dot{o}_{i,\pm}(\varepsilon^{1+l_{i,\pm}})$, $\varepsilon \in (0, \varepsilon_0]$, for $i \in \mathbb{X}$. According to conditions **D**

and **E**, these asymptotic expansions have non-zero first coefficients. Therefore, all asymptotic expansions penetrating conditions **D** and **E** have the length 2. As we shall see, this makes it possible to represent the stationary and conditional quasi-stationary probabilities in the form of asymptotic expansions of the length 2.

Note that conditions **D** and **E** imply that there exists $\varepsilon_0' \in (0, \varepsilon_0]$ such that the probabilities $p_{i,\pm}(\varepsilon) > 0, i \in \mathbb{X}$ and the expectations $e_{i,\pm}(\varepsilon) > 0, i \in \mathbb{X}$ for $\varepsilon \in (0, \varepsilon_0']$. Therefore, let us just assume that $\varepsilon_0' = \varepsilon_0$.

The model assumption, $p_{i,-}(\varepsilon) + p_{i,+}(\varepsilon) = 1, \varepsilon \in (0, \varepsilon_0]$, also implies that the following condition should hold:

F: $a_{i,-}[0] + a_{i,+}[0] = 1, a_{i,-}[1] + a_{i,+}[1] = 0$, for $i \in \mathbb{X}$.

We also assume that the following natural consistency condition for asymptotic expansions penetrating perturbation conditions **D** and **E** holds:

G: $b_{i,\pm}[0] > 0$ if and only if $a_{i,\pm}[0] > 0$, for $i = 0, N$.

There are three basic variants of the model that correspond to (11.7) and (11.8). For the more general setup of semi-Markov chains in this section, we formulate this a bit differently and assume that one of the following conditions holds:

$\mathbf{H_1}$: $a_{0,+}[0] > 0, a_{N,-}[0] > 0$.

$\mathbf{H_2}$: $a_{0,+}[0] = 0, a_{N,-}[0] > 0$.

$\mathbf{H_3}$: $a_{0,+}[0] = 0, a_{N,-}[0] = 0$.

The case $a_{0,+}[0] > 0, a_{N,-}[0] = 0$ is analogous to the case where condition $\mathbf{H_2}$ holds and we omit its consideration.

Condition **D** implies that there exist $\lim_{\varepsilon \to 0} p_{i,\pm}(\varepsilon) = p_{i,\pm}(0), i \in \mathbb{X}$ and, thus, there also exist $\lim_{\varepsilon \to 0} p_{ij}(\varepsilon) = p_{ij}(0), i, j \in \mathbb{X}$. Condition **E** implies that there exist $\lim_{\varepsilon \to 0} e_{i,\pm}(\varepsilon) = e_{i,\pm}(0), i \in \mathbb{X}$ and, thus, there also exist $\lim_{\varepsilon \to 0} e_{ij}(\varepsilon) = e_{ij}(0), i, j \in \mathbb{X}$.

The limiting birth-death type Markov chain $\eta_n^{(0)}$ with the matrix of transition probabilities $\|p_{ij}(0)\|$ has: **(a)** one class of communicative states $\mathbb{X}$, if condition $\mathbf{H_1}$ holds, **(b)** one communicative class of transient states ${}_{\langle 1,N \rangle}\mathbb{X} = \mathbb{X} \setminus \{0\}$ and an absorbing state 0, if condition $\mathbf{H_2}$ holds, and **(c)** one communicative class of transient states ${}_{\langle 1,N-1 \rangle}\mathbb{X} = \mathbb{X} \setminus \{0, N\}$ and two absorbing states 0 and N, if condition $\mathbf{H_3}$ holds.

In this paper, we get, under conditions **A**–**G** and $\mathbf{H}_i$ (for $i = 1, 2, 3$), asymptotic expansions for stationary probabilities, as $\varepsilon \to 0$,

$$\pi_i(\varepsilon) = \sum_{l=\dot{l}_i}^{1+\dot{l}_i} c_i[l]\varepsilon^l + o_i(\varepsilon^{1+\dot{l}_i}), \ i \in \mathbb{X}, \tag{11.36}$$

where: (a) $\dot{l}_i = 0, i \in \mathbb{X}$ and the limiting stationary probabilities $\pi_i(0) > 0, i \in \mathbb{X}$, if condition $\mathbf{H_1}$ holds, (b) $\dot{l}_i = I(i \neq 0), i \in \mathbb{X}$ and $\pi_0(0) = 1, \pi_i(0) = 0, i \in$

$_{\langle 1,N\rangle}\mathbb{X}$, if condition $\mathbf{H_2}$ holds, and (c) $\dot{l}_i = I(i \neq 0, N)$, $i \in \mathbb{X}$ and $\pi_0(0), \pi_N(0) > 0$, $\pi_0(0) + \pi_N(0) = 1$, $\pi_i(0) = 0$, $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$, if condition $\mathbf{H_3}$ holds.

This implies that there is sense to consider so-called conditional quasi-stationary probabilities, which are defined as,

$$\tilde{\pi}_i(\varepsilon) = \frac{\pi_i(\varepsilon)}{1 - \pi_0(\varepsilon)} = \frac{\pi_i(\varepsilon)}{\sum_{j \in {}_0\mathbb{X}} \pi_j(\varepsilon)}, \ i \in {}_{\langle 1,N\rangle}\mathbb{X}, \tag{11.37}$$

in the case where condition $\mathbf{H_2}$ holds, or as,

$$\hat{\pi}_i(\varepsilon) = \frac{\pi_i(\varepsilon)}{1 - \pi_0(\varepsilon) - \pi_N(\varepsilon)} = \frac{\pi_i(\varepsilon)}{\sum_{j \in {}_{0,N}\mathbb{X}} \pi_j(\varepsilon)}, \ \ i \in {}_{\langle 1,N-1\rangle}\mathbb{X}, \tag{11.38}$$

in the case where condition $\mathbf{H_3}$ holds.

We also get, under conditions **A**–**G** and $\mathbf{H_2}$, asymptotic expansions for conditional quasi-stationary probabilities,

$$\tilde{\pi}_i(\varepsilon) = \sum_{l=0}^{1} \tilde{c}_i[l]\varepsilon^l + \tilde{o}_i(\varepsilon), \ i \in {}_{\langle 1,N\rangle}\mathbb{X}, \tag{11.39}$$

and, under conditions **A**–**G** and $\mathbf{H_3}$, asymptotic expansions for conditional quasi-stationary probabilities,

$$\hat{\pi}_i(\varepsilon) = \sum_{l=0}^{1} \hat{c}_i[l]\varepsilon^l + \hat{o}_i(\varepsilon), \ i \in {}_{\langle 1,N-1\rangle}\mathbb{X}. \tag{11.40}$$

The coefficients in the above asymptotic expansions are given by explicit formulas via coefficients in asymptotic expansions given in initial perturbation conditions **D** and **E**.

As it was mentioned in the introduction, the first coefficients $\pi_i(\varepsilon) = c_i[0]$, $\tilde{\pi}_i(0) = \tilde{c}_i[0]$ and $\hat{\pi}_i(0) = \hat{c}_i[0]$ describe the asymptotic behaviour of stationary and quasi-stationary probabilities and their continuity properties with respect to small perturbations of transition characteristics of the corresponding semi-Markov processes. The second coefficients $c_i[1]$, $\tilde{c}_i[1]$ and $\hat{c}_i[1]$ determine sensitivity of stationary and quasi-stationary probabilities with respect to small perturbations of transition characteristics.

We also would like to comment the use of the term "conditional quasi-stationary probability" for quantities defined in relations (11.37) and (11.38). As a matter of fact, the term "quasi-stationary probability (distribution)" is traditionally used for limits,

$$q_j(\varepsilon) = \lim_{t \to \infty} \mathsf{P}_i\{\eta^{(\varepsilon)}(t) = j / \eta^{(\varepsilon)}(s) \notin A, 0 \leq s \leq t\}, \tag{11.41}$$

where A is some special subset of $\mathbb{X}$.

A detailed presentation of results concerning quasi-stationary distributions and comprehensive bibliographies of works in this area can be found in the books by Gyllenberg and Silvestrov [13], Nåsell [34] and Collet, Martínez and San Martín [6]. We would also like to mention the paper by Allen and Tarnita [1], where one can find a discussion concerning the above two forms of quasi-stationary distributions for some bio-stochastic systems.

11.4 Examples of Stationary Distributions

In this section, we will revisit the examples of Sect. 11.2 and illustrate how to compute, approximate and expand the various stationary and conditional quasi-stationary distributions that were introduced in Sect. 11.3. Since all the models of Sect. 11.2 have a geometric or exponential transition time distribution (11.3)–(11.4), and since the transition probabilities satisfy (11.28), it follows that the stationary distribution (11.34)–(11.35) has a very explicit expression,

$$\pi_i(\varepsilon) \propto \begin{cases} 1, & i = 0, \\ \frac{\lambda_{0,+}(\varepsilon)\cdot\ldots\cdot\lambda_{i-1,+}(\varepsilon)}{\lambda_{1,-}(\varepsilon)\cdot\ldots\cdot\lambda_{i,-}(\varepsilon)}, & i = 1, \ldots, N, \end{cases} \tag{11.42}$$

for $0 < \varepsilon \leq \varepsilon_0$, with a proportionality constant chosen so that $\sum_{i=0}^{N} \pi_i(\varepsilon) = 1$.

Our goal is to find a series representation of (11.42). Since the models of Sect. 11.2 are formulated in terms of the death and birth rates in (11.5), we will assume that these rates admit expansions

$$\lambda_{i,\pm}(\varepsilon) = \sum_{l=0}^{L_{i,\pm}} g_{i,\pm}[l]\varepsilon^l + o_{i,\pm}(\varepsilon^{L+l_{i,\pm}}) \tag{11.43}$$

for $\varepsilon \in (0, \varepsilon_0]$, and then check the regularity conditions of Sect. 11.3 that are needed to hold. From Eqs. (11.3)–(11.4), (11.6), and (11.31), we deduce that

$$e_{i,\pm}(\varepsilon) = \frac{1}{\lambda_i(\varepsilon)} \cdot \frac{\lambda_{i,\pm}(\varepsilon)}{\lambda_i(\varepsilon)}. \tag{11.44}$$

Inserting (11.43) into (11.44), we find that

$$g_{i,-}[0] + g_{i,+}[0] > 0 \tag{11.45}$$

must hold for all $i \in \mathbb{X}$ in order for the series expansion of $e_{i,\pm}(\varepsilon)$ to satisfy condition $\mathbf{E}_L$. It therefore follows from (11.6) that $p_{i,\pm}(\varepsilon)$ will satisfy perturbation condition $\mathbf{D}_L$, with $L + l_{i,+} = L + l_{i,-} = \min(L_{i,-}, L_{i,+})$, and

$$a_{i,\pm}[0] = \frac{g_{i,\pm}[0]}{g_{i,-}[0] + g_{i,+}[0]}. \tag{11.46}$$

Because of (11.45) and (11.46), we can rephrase the three perturbation scenarios $\mathbf{H}_1$ – $\mathbf{H}_3$ of Sect. 11.3.2 as

$$\begin{aligned} \mathbf{H}_1 &: g_{0,+}[0] > 0,\, g_{N,-}[0] > 0, \\ \mathbf{H}_2 &: g_{0,+}[0] = 0,\, g_{N,-}[0] > 0, \\ \mathbf{H}_3 &: g_{0,+}[0] = 0,\, g_{N,-}[0] = 0, \end{aligned} \tag{11.47}$$

in agreement with (11.8). Under $\mathbf{H}_2$, the exact expression for the conditional quasi-stationary distribution (11.37) is readily obtained from (11.42). It equals

$$\tilde{\pi}_i(\varepsilon) \propto \frac{\lambda_{1,+}(\varepsilon) \cdot \ldots \cdot \lambda_{i-1,+}(\varepsilon)}{\lambda_{1,-}(\varepsilon) \cdot \ldots \cdot \lambda_{i,-}(\varepsilon)} \tag{11.48}$$

for $i \in {}_0\mathbb{X}$ and $0 < \varepsilon \leq \varepsilon_0$, with the numerator equal to 1 when $i = 1$, and a proportionality constant chosen so that $\sum_{i=1}^{N} \tilde{\pi}_i(\varepsilon) = 1$. As $\varepsilon \to 0$, this expression converges to

$$\tilde{\pi}_i(0) \propto \frac{\lambda_{1,+}(0) \cdot \ldots \cdot \lambda_{i-1,+}(0)}{\lambda_{1,-}(0) \cdot \ldots \cdot \lambda_{i,-}(0)}. \tag{11.49}$$

If scenario $\mathbf{H}_3$ holds, we find analogously that the conditional quasi-stationary distribution (11.38) is given by

$$\hat{\pi}_i(\varepsilon) \propto \frac{\lambda_{1,+}(\varepsilon) \cdot \ldots \cdot \lambda_{i-1,+}(\varepsilon)}{\lambda_{1,-}(\varepsilon) \cdot \ldots \cdot \lambda_{i,-}(\varepsilon)} \tag{11.50}$$

for $i \in_{0,N} \mathbb{X}$, with a limit

$$\hat{\pi}_i(0) \propto \frac{\lambda_{1,+}(0) \cdot \ldots \cdot \lambda_{i-1,+}(0)}{\lambda_{1,-}(0) \cdot \ldots \cdot \lambda_{i,-}(0)}. \tag{11.51}$$

11.4.1 Stationary Distributions for Perturbed Population Dynamics Models

For the population dynamics model (11.9) of Sect. 11.2.1, we considered two perturbation scenarios. Recall that the first one in (11.12) has a varying immigration parameter $\nu(\varepsilon) = \varepsilon$, whereas all other parameters are kept fixed. Since $\lambda_{0,-}(\varepsilon) = 0$ and $\lambda_{0,+}(\varepsilon) = \varepsilon$, it follows that $g_{0,-}[0] = g_{0,+}[0] = 0$, and therefore formula (11.45) is violated for $i = 0$. But the properties of $\eta^{(\varepsilon)}$ remain the same if we put $\lambda_{0,-}(\varepsilon) = 1$ instead. With this modification, formula (11.47) implies that condition $\mathbf{H}_2$ of

Sect. 11.3.2 holds, and hence the $\varepsilon \to 0$ limit of the stationary distribution in (11.34) and (11.42) is concentrated at state 0 ($\pi_0(0) = 1$).

Let $\tau_0^{(\varepsilon)}$ be the time it takes for the population to get temporarily extinct again, after an immigrant has entered an empty island. It then follows from a slight modification of Eq. (11.90) in Sect. 11.5.2 and the relation $\lambda_{0,+}(\varepsilon) = \varepsilon$, that a first order expansion of the probability that the island is empty at stationarity, is

$$\pi_0(\varepsilon) = \frac{1/\lambda_{0,+}(\varepsilon)}{1/\lambda_{0,+}(\varepsilon) + \mathsf{E}_1(\tau_0^{(\varepsilon)})} = \frac{1/\varepsilon}{1/\varepsilon + E_{10}(\varepsilon)} = 1 - E_{10}(\varepsilon)\varepsilon + o(\varepsilon). \quad (11.52)$$

This expansion is accurate when the perturbation parameter is small ($\varepsilon \ll 1/E_{10}(\varepsilon)$), otherwise higher order terms in (11.52) are needed. The value of $E_{10}(\varepsilon)$ will be highly dependent on the value of the basic reproduction number R_0 in (11.13). When $R_0 > 1$, the expected time to extinction will be very large, and $\pi_0(\varepsilon)$ will be close to 0 for all but very small ε. On the other hand, (11.52) is accurate for a larger range of ε when $R_0 < 1$, since $E_{10}(\varepsilon)$ is then small.

In order to find useful approximations of the conditional quasi-stationary distribution $\tilde{\pi}_i(\varepsilon)$ in (11.48), we will distinguish between whether R_0 is larger than or smaller than 1. When $R_0 > 1$, or equivalently $\lambda > \mu$, we can rewrite (11.11) as

$$\mathsf{E}\left[\eta^{(\varepsilon)}(t + \Delta t) - \eta^{(\varepsilon)}(t) | \eta^{(\varepsilon)}(t) = i\right] = \Delta t \cdot Nm\left(\frac{i}{N}\right), \quad (11.53)$$

where

$$m(x) = rx + \frac{\varepsilon}{N} - \left[rx(0)^{-\theta} \cdot x + \frac{\varepsilon}{N}\right] x^{\theta} \quad (11.54)$$

is a rescaled mean function of the drift, $r = \mu(R_0 - 1)$ is the intrinsic growth rate, or growth rate per capita, of a small population without immigration ($\varepsilon = 0$), and

$$x(0) = \left(\frac{R_0 - 1}{\alpha_1 R_0 + \alpha_2}\right)^{1/\theta}.$$

We assume that α_1 and α_2 are large enough so that $x(0) < 1$. A sufficient condition for this is $\alpha_1 + \alpha_2 = 1$. The carrying capacity $K(\varepsilon) = Nx(\varepsilon)$ of the environment is the value of i such that the right hand side of (11.53) equals zero. We can write $x = x(\varepsilon)$ as the unique solution of $m(x) = 0$, or equivalently

$$x^{\theta} = \frac{rx + \varepsilon N^{-1}}{rx(0)^{-\theta}x + \varepsilon N^{-1}},$$

with $x(\varepsilon) \searrow x(0)$ as $\varepsilon \to 0$. The conditional quasi-stationary distribution (11.48) will be centred around $K(\varepsilon)$. In order to find a good approximation of this distribution, we look at the second moment

$$\begin{aligned}\mathsf{E}\left\{\left[\eta^{(\varepsilon)}(t+\Delta t)-\eta^{(\varepsilon)}(t)\right]^2|\eta^{(\varepsilon)}(t)=i\right\}&=\Delta t\left[\lambda_{i,+}(\varepsilon)+\lambda_{i,-}(\varepsilon)\right]\\&=\Delta t\cdot Nv\left(\tfrac{i}{N}\right),\end{aligned}$$

of the drift of $\eta^{(\varepsilon)}$, with

$$v(x)=\lambda x(1-\alpha_1 x^\theta)+\frac{\varepsilon}{N}(1-x^\theta)+\mu x(1+\alpha_2 x^\theta). \tag{11.55}$$

When N is large, we may approximate the conditional quasi-stationary distribution

$$\begin{aligned}\tilde{\pi}_i(\varepsilon)&\approx\textstyle\int_{i-}^{i+}f^{(\varepsilon)}(k)dk\\&=\textstyle\int_{i-}^{i+}f^{(0)}(k)dk+\int_{i-}^{i+}\left.\frac{df^{(\varepsilon)}(k)}{d\varepsilon}\right|_{\varepsilon=0}dk\cdot\varepsilon+o(\varepsilon),\end{aligned} \tag{11.56}$$

by integrating a density function $f^{(\varepsilon)}$ on $[0,N]$ between $i_-=\max(0,i-1/2)$ and $i_+=\min(N,i+1/2)$. This density function can be found through a diffusion argument as the stationary density

$$\begin{aligned}f^{(\varepsilon)}(k)&\propto\tfrac{1}{Nv(\frac{k}{N})}\exp\left(2\textstyle\int_{K(\varepsilon)}^{k}\frac{Nm(\frac{y}{N})}{Nv(\frac{y}{N})}dy\right)\\&\propto\tfrac{1}{v(\frac{k}{N})}\exp\left(2\textstyle\int_{K(\varepsilon)}^{k}\frac{m(\frac{y}{N})}{v(\frac{y}{N})}dy\right)\end{aligned} \tag{11.57}$$

of Kolmogorov's forward equation, with a proportionality constant chosen so that $\int_0^N f^{(\varepsilon)}(k)dk=1$ (see, for instance, Chap. 9 of Crow and Kimura [7]). A substitution of variables $x=y/N$ in (11.57), and a Taylor expansion of $m(x)$ around $x(\varepsilon)$ reveals that the diffusion density has approximately a normal distribution

$$f^{(\varepsilon)}\sim N\left(K(\varepsilon),N\frac{v\left[x(\varepsilon)\right]}{2|m'\left[x(\varepsilon)\right]|}\right). \tag{11.58}$$

Expansion (11.56) is valid for small migration rates ε, and its linear term quantifies how sensitive the conditional quasi-stationary distribution is to a small amount of immigration.

It follows from (11.53) that the expected population size

$$\mathsf{E}\left[\eta^{(0)}(t+\Delta t)-\eta^{(0)}(t)|\eta^{(0)}(t)=i\right]=\Delta t\cdot ri\left[1-\left(\frac{i}{K(0)}\right)^\theta\right] \tag{11.59}$$

of an isolated population varies according to a theta logistic model (Gilpin and Ayala [12]), which is a special case of the generalised growth curve model in Tsoularis and Wallace [44]. The theta logistic model has a carrying capacity $K(0)$ of the environment to accommodate new births. When $\theta=1$, we obtain the logistic growth model of Verhulst [45]. Pearl [35] used such a curve to approximate population growth in the United States, and Feller [10] introduced a stochastic version of the

logistic model in terms of a Markov birth-death process (11.4) in continuous time. Feller's approach has been extended for instance by Kendall [21], Whittle [47], and Nåsell [31, 33]. In particular, Nåsell studied the quasi-stationary distribution (11.41) of $\eta^{(\varepsilon)}$, with $A = \{1, \ldots, N\}$. In this paper the previously studied population growth models are generalised in two directions; we consider semi-Markov processes and allow for theta logistic expected growth.

When $0 < R_0 < 1$, or equivalently $0 < \lambda < \mu$, we rewrite (11.11) as

$$\begin{aligned} &\mathsf{E}\left[\eta^{(\varepsilon)}(t + \Delta t) - \eta^{(\varepsilon)}(t) | \eta^{(\varepsilon)}(t) = i\right] \\ &\quad = \Delta t \cdot \left[\nu - ri - (\nu + ri\tilde{x}^{-\theta})\left(\tfrac{i}{N}\right)^{\theta}\right], \end{aligned} \tag{11.60}$$

where $r = (1 - R_0)\mu$ quantifies per capita decrease for a small population without immigration, and $\tilde{x} = [(1 - R_0)/(\alpha_1 R_0 + \alpha_2)]^{1/\theta}$ is the fraction of the maximal population size at which the per capita decrease of an isolated $\varepsilon = 0$ population has doubled to $2r$. For large N, we can neglect all $O(N^{-\theta})$ terms, and it follows from (11.49) that

$$\tilde{\pi}_i(\varepsilon) \approx \frac{1}{\log(1 - R_0)} \cdot \frac{R_0^i}{i} + \tilde{c}_i[1]\varepsilon + o(\varepsilon),$$

for $i = 1, \ldots, N$.

Recall that the second perturbation scenario (11.14) has a varying birth rate $\langle(\varepsilon) = \varepsilon$, whereas all other parameters are kept fixed, not depending on ε. In view of (11.47), it satisfies condition $\mathbf{H}_1$ of Sect. 11.3.2. Suppose N is large. If $\nu = o(N)$, it follows from (11.42) that the stationary distribution for small values of ε is well approximated by

$$\pi_i(\varepsilon) \approx \frac{(\nu/\mu)^i}{i!} e^{-\nu/\mu} + c_i[1]\varepsilon + o(\varepsilon)$$

for $i = 0, \ldots, N$, a Poisson distribution with mean ν/μ, corrupted by a sensitivity term $c_i[1]\varepsilon$ due to births. If $\nu = VN$, the carrying capacity of the environment is $K(\varepsilon) = Nx(\varepsilon)$, where $x = x(\varepsilon)$ is the value of i/N in (11.60) such that the right hand side vanishes, i.e. the unique solution of the equation

$$rx + Vx^{\theta} + r\tilde{x}^{-\theta}x^{\theta+1} = V,$$

with $r = r(\varepsilon) = \mu - \varepsilon$. The stationary distribution (11.42) is well approximated by a discretised normal distribution (11.56)–(11.58), but with a mean drift function $m(x)$ obtained from (11.60), and a variance function $v(x)$ derived similarly.

11.4.2 Stationary Distributions for Perturbed Epidemic Models

For the epidemic models of Sect. 11.2.2, we considered one perturbation scenario (11.18), with a varying external contact rate $\nu(\varepsilon) = \varepsilon$. When the basic reproduction model $R_0 = \mu/\mu$ exceeds one, the expected growth rate follows a logistic model (11.18) when $\varepsilon = 0$, which is a special case of the theta logistic mean growth curve model (11.59), with $\theta = 1$. When $R_0 < 1$, we similarly write the expected population decline as in (11.60), with $\theta = 1$. Since the SIS model is a particular case of the population dynamic models of Sect. 11.2.1 (Nåsell, [34]), the stationary and conditional quasi-stationary distributions are approximated in the same way as in Sect. 11.4.1.

11.4.3 Stationary Distributions for Perturbed Models of Population Genetics

For the population genetics model of Sect. 11.2.3, we recall there were three different perturbation scenarios (11.27). For all of them, the rescaled mutation rates $U_1(\varepsilon) = N\mathsf{P}(A_1 \to A_2)$ and $U_2(\varepsilon) = N\mathsf{P}(A_2 \to A_1)$ between the two alleles A_1 and A_2 are linear functions of ε.

The stationary distribution is either found by first inserting (11.20) into (11.5), and then (11.5) into (11.42), or, for large N, it is often more convenient to use a diffusion approximation,

$$\pi_i(\varepsilon) \approx \int_{x_{i,-}}^{x_{i,+}} f^{(\varepsilon)}(x)dx. \tag{11.61}$$

It is obtained by integrating the density function

$$\begin{aligned} f^{(\varepsilon)}(x) &\propto \tfrac{1}{\nu(x)} \exp\left(2\textstyle\int_{1/2}^{x} \frac{m(y)}{\nu(y)}dy\right) \\ &\propto (1-x)^{-1+U_1} x^{-1+U_2} \exp\left[\tfrac{1}{2}(S_1+S_2)x^2 - S_2 x\right] \end{aligned} \tag{11.62}$$

between $x_{i,-} = \max[0, (i-1/2)/N]$ and $x_{i,+} = \min[1, (i+1/2)/N]$. This density is defined in terms of the infinitesimal drift and variance functions $m(x)$ and $\nu(x)$ in (11.23)–(11.25), with a constant of proportionality chosen to ensure that $\int f^{(\varepsilon)}(x)dx = 1$. See, for instance, Chap. 9 of Crow and Kimura [7] and Chap. 7 of Durrett [8] for details.

For $\mathbf{H}_1$, we use this diffusion argument to find an approximate first order series expansion

$$\pi_i(\varepsilon) \approx \int_{x_{i,-}}^{x_{i,+}} f^{(0)}(x)dx + \int_{x_{i,-}}^{x_{i,+}} \left.\frac{df^{(\varepsilon)}(x)}{d\varepsilon}\right|_{\varepsilon=0} dx \cdot \varepsilon + o(\varepsilon)$$

of the stationary distribution by inserting (11.26) into (11.61)–(11.62). The null density $f^{(0)}(x)$ is defined by (11.62), with C_1 and C_2 instead of U_1 and U_2. For a neutral model ($S_1 = S_2 = 0$), the stationary null distribution is approximately beta with parameters C_1 and C_2, and expected value $C_2/(C_1 + C_2)$. A model with $S_1 > 0 > S_2$ corresponds to directional selection, with higher fitness for A_1 compared to A_2. It can be seen from (11.62) that the stationary null distribution is further skewed to the right than for a neutral model. A model with balancing selection or overdominance has negative S_1 and S_2, so that the heterozygous genotype A_1A_2 has a selective advantage. The stationary null distribution will then have a peak around $S_2/(S_1 + S_2)$. On the other hand, for an underdominant model where S_1 and S_2 are both positive, the heterozygous genotype will have a selective disadvantage. Then $S_2/(S_1 + S_2)$ functions as a repelling point of the stationary null distribution.

For scenario $\mathbf{H}_2$, the null model has one absorbing state 0. In analogy with (11.52), we find that the series expansion of the stationary probability of no A_1 alleles in the population is

$$\pi_0(\varepsilon) = 1 - \mathsf{E}_1(\tau_0^{(\varepsilon)}) \cdot \frac{D_2\varepsilon}{N} + o(\varepsilon)$$

when $D_2 > 0$, for small values of the perturbation parameter. Here $D_2\varepsilon/N$ is the probability that a mutation $A_2 \to A_1$ occurs in a homogeneous A_2 population, and $\tau_0^{(\varepsilon)}$ is the time it takes for the A_1 allele to disappear again.

Because of the singularity at $i = 0$ for small ε, we avoid the diffusion argument and find the conditional quasi-stationary distribution (11.37) directly by first inserting (11.20) into (11.5), and then (11.5) into (11.48)–(11.49). After some computations, this leads to

$$\begin{aligned}\tilde{\pi}_i(\varepsilon) \approx{}& \tilde{c}_1[0] i^{-1} \left(1 - \tfrac{i-1}{N}\right)^{C_1 - 1} \exp\left[\tfrac{1}{2}(S_1 + S_2)\tfrac{i-1}{N}\tfrac{i}{N} - S_2\tfrac{i-1}{N}\right] \\ &+ \tilde{c}_1[1]\varepsilon + o(\varepsilon)\end{aligned} \tag{11.63}$$

for $i = 1, \ldots, N$, where $\tilde{c}_1[0]$ is chosen so that $\sum_{i=1}^{N} \tilde{\pi}_i(0) = 1$, and $\tilde{c}_1[1]$ will additionally involve D_1 and D_2. If $D_2 = 0$, we have that $\pi_0(\varepsilon) = 1$ for all $0 < \varepsilon \leq \varepsilon_0$, so that the conditional quasi-stationary distribution (11.37) is not well defined. However, the time to reach absorption is very large for small $U_1 > 0$. It is shown in Hössjer, Tyvand and Miloh [17] that $\eta^{(\varepsilon)}$ may be quasi-fixed for a long time at the other boundary point $i = N$, before eventual absorption at $i = 0$ occurs.

For scenario $\mathbf{H}_3$, the null model is mutation free, and the asymptotic distribution

$$P_j(0; i) = \lim_{t \to \infty} \mathsf{P}_i(\eta^{(0)}(t) = j)$$

is supported on the two absorbing states ($j \in \{0, N\}$), and it is dependent on the state i at which the process starts. For a neutral model ($s_1 = s_2 = 0$), we have that

$$P_N(0; i) = 1 - P_0(0; i) = \frac{i}{N}. \tag{11.64}$$

A particular case of directional selection is multiplicative fitness, with $1 + s_1 = (1 + s_2)^{-1}$. It is mathematically simpler since selection operates directly on alleles, not on genotypes, with selective advantages 1 and $1 + s_2$ for A_1 and A_2. It follows for instance from Sect. 6.1 of Durrett [8] that

$$P_N(0; i) = 1 - P_0(0; i) = \frac{1 - (1 + s_2)^i}{1 - (1 + s_2)^N} \tag{11.65}$$

for multiplicative fitness. Notice that $P_0(0; i)$ and $P_N(0; i)$ will differ from $\pi_j(0) = \lim_{\varepsilon \to 0} \pi_0(\varepsilon)$ at the two boundaries. Indeed, by ergodicity (11.34) for each $\varepsilon > 0$, the latter two probabilities are not functions of $i = \eta^{(0)}(0)$. From (11.61)–(11.62), we find that

$$\pi_N(0) = 1 - \pi_0(0) \approx \frac{D_2}{\exp\left[-\frac{1}{2}(S_1 - S_2)\right] D_1 + D_2}. \tag{11.66}$$

Similarly as in (11.63), we find after some computations that the conditional quasi-stationary distribution (11.38) admits an approximate expansion

$$\begin{aligned} \hat{\pi}_i(\varepsilon) \approx{}& \hat{c}_1[0] i^{-1} \left(1 - \tfrac{i-1}{N}\right)^{-1} \exp\left[\tfrac{1}{2}(S_1 + S_2)\tfrac{i-1}{N}\tfrac{i}{N} - S_2 \tfrac{i-1}{N}\right] \\ &+ \hat{c}_1[1]\varepsilon + o(\varepsilon) \end{aligned} \tag{11.67}$$

for $i = 1, \ldots, N - 1$, where $\hat{c}_1[0]$ is chosen so that $\sum_{i=1}^{N-1} \hat{\pi}_i(0) = 1$, and $\hat{c}_1[1]$ will additionally involve D_1 and D_2. Notice that the limiting fixation probabilities in (11.66) are functions of the mutation probability ratio $D_1/(D_1 + D_2)$, but the limiting conditional quasi-stationary distribution $\hat{\pi}_i(0)$ in (11.67) does not involve any of D_1 or D_2.

11.5 Reduced Semi-Markov Birth-Death Processes

In this section, we present a time-space screening procedure of phase space reduction for perturbed semi-Markov birth-death processes and recurrent algorithms for computing expectations of hitting times and stationary and conditional quasi-stationary distributions for such processes.

11.5.1 Phase Space Reduction for Semi-Markov Birth-Death Processes

Let us assume that $N \geq 1$. Let $0 \leq k \leq i \leq r \leq N$ and define the reduced phase space ${}_{\langle k,r \rangle}\mathbb{X} = \{k, \ldots, r\}$. Note that, by the definition, ${}_{\langle 0,N \rangle}\mathbb{X} = \mathbb{X}$. Let us also assume

that the initial distribution $\bar{p}^{(\varepsilon)}$ is concentrated on the phase space ${}_{\langle k,r\rangle}\mathbb{X}$, i.e. $p_i^{(\varepsilon)} = 0, i \notin {}_{\langle k,r\rangle}\mathbb{X}$.

Let us define the sequential moments of hitting the reduced space ${}_{\langle k,r\rangle}\mathbb{X}$, by the embedded Markov chain $\eta_n^{(\varepsilon)}$,

$$ {}_{\langle k,r\rangle}\xi_n^{(\varepsilon)} = \min(k > {}_{\langle k,r\rangle}\xi_{n-1}^{(\varepsilon)},\ \eta_k^{(\varepsilon)} \in {}_{\langle k,r\rangle}\mathbb{X}),\ n \geq 1,\ {}_{\langle k,r\rangle}\xi_0^{(\varepsilon)} = 0. \tag{11.68} $$

Now, let us define the random sequence,

$$ ({}_{\langle k,r\rangle}\eta_n^{(\varepsilon)},\ {}_{\langle k,r\rangle}\kappa_n^{(\varepsilon)}) = \begin{cases} (\eta_0^{(\varepsilon)}, 0) & \text{for } n = 0, \\ (\eta^{(\varepsilon)}_{{}_{\langle k,r\rangle}\xi_n^{(\varepsilon)}}, \sum_{l={}_{\langle k,r\rangle}\xi_{n-1}^{(\varepsilon)}+1}^{{}_{\langle k,r\rangle}\xi_n^{(\varepsilon)}} \kappa_l^{(\varepsilon)}) & \text{for } n \geq 1. \end{cases} \tag{11.69} $$

This sequence is a Markov renewal process with a phase space ${}_{\langle k,r\rangle}\mathbb{X} \times [0,\infty)$, the initial distribution $\bar{p}^{(\varepsilon)}$, and transition probabilities defined for $(i,s),(j,t) \in \mathbb{X} \times [0,\infty)$,

$$ {}_{\langle k,r\rangle}Q_{ij}^{(\varepsilon)}(t) = \mathsf{P}\{ {}_{\langle k,r\rangle}\eta_1^{(\varepsilon)} = j,\ {}_{\langle k,r\rangle}\kappa_1^{(\varepsilon)} \leq t / {}_{\langle k,r\rangle}\eta_0^{(\varepsilon)} = i,\ {}_{\langle k,r\rangle}\kappa_0^{(\varepsilon)} = s\}. \tag{11.70} $$

We define a reduced semi-Markov process by

$$ {}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t) = {}_{\langle k,r\rangle}\eta^{(\varepsilon)}_{{}_{\langle k,r\rangle}\nu^{(\varepsilon)}(t)},\ t \geq 0, \tag{11.71} $$

where ${}_{\langle k,r\rangle}\nu^{(\varepsilon)}(t) = \max(n \geq 0 :\ {}_{\langle k,r\rangle}\zeta_n^{(\varepsilon)} \leq t)$ is the number of jumps in the time interval $[0,t]$, for $t \geq 0$, and ${}_{\langle k,r\rangle}\zeta_n^{(\varepsilon)} = {}_{\langle k,r\rangle}\kappa_1^{(\varepsilon)} + \cdots + {}_{\langle k,r\rangle}\kappa_n^{(\varepsilon)}$, $n = 0,1,\ldots$ are sequential moments of jumps, for the semi-Markov process ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$.

In particular, the initial semi-Markov process $\eta^{(\varepsilon)}(t) = {}_{\langle 0,N\rangle}\eta^{(\varepsilon)}(t)$.

It is readily seen that ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$ is also a semi-Markov birth-death process, i.e. the time-space screening procedure of phase space reduction described above preserves the birth-death structure of the semi-Markov birth-death process $\eta^{(\varepsilon)}(t)$.

11.5.2 Expectations of Hitting Times for Reduced Semi-Markov Birth-Death Processes

Let us now introduce hitting times for semi-Markov birth-death process $\eta^{(\varepsilon)}(t)$. We define hitting times, which are random variables given by the following relation, for $j \in \mathbb{X}$,

$$ \tau_j^{(\varepsilon)} = \sum_{n=1}^{\nu_j^{(\varepsilon)}} \kappa_n^{(\varepsilon)}, \tag{11.72} $$

where $\nu_j^{(\varepsilon)} = \min(n \geq 1 : \eta_n^{(\varepsilon)} = j)$.

Let us denote,

$$E_{ij}(\varepsilon) = \mathsf{E}_i \tau_j^{(\varepsilon)},\ i, j \in \mathbb{X}. \tag{11.73}$$

As is known, conditions **A**–**C** imply that, for every $\varepsilon \in (0, \varepsilon_0]$, expectations of hitting times are finite, i.e,

$$0 < E_{ij}(\varepsilon) < \infty,\ i, j \in \mathbb{X}. \tag{11.74}$$

We also denote by ${}_{\langle k,r\rangle}\tau_j^{(\varepsilon)}$ the hitting time to the state $j \in {}_{\langle k,r\rangle}\mathbb{X}$ for the reduced semi-Markov birth-death process ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$.

The following theorem, which proof can be found, for example, in Silvestrov and Manca [38], plays the key role in what follows.

Theorem 11.1 *Let conditions* **A**–**C** *hold for the semi-Markov birth-death process* $\eta^{(\varepsilon)}(t)$. *Then, for any state* $j \in {}_{\langle k,r\rangle}\mathbb{X}$, *the first hitting times* $\tau_j^{(\varepsilon)}$ *and* ${}_{\langle k,r\rangle}\tau_j^{(\varepsilon)}$ *to the state* j, *respectively, for semi-Markov processes* $\eta^{(\varepsilon)}(t)$ *and* ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$, *coincide, and, thus, the expectations of hitting times* $E_{ij}(\varepsilon) = \mathsf{E}_i \tau_j^{(\varepsilon)} = \mathsf{E}_i\, {}_{\langle k,r\rangle}\tau_j^{(\varepsilon)}$, *for any* $i, j \in {}_{\langle k,r\rangle}\mathbb{X}$ *and* $\varepsilon \in (0, \varepsilon_0]$.

11.5.3 Sequential Reduction of Phase Space for Semi-Markov Birth-Death Processes

Let us consider the case, where the left end state 0 is excluded from the phase space $\mathbb{X}$. In this case, the reduced phase space ${}_{\langle 1,N\rangle}\mathbb{X} = \{1, \ldots, N\}$.

We assume that the initial distribution of the semi-Markov process $\eta^{(\varepsilon)}(t)$ is concentrated on the reduced phase space ${}_{\langle 1,N\rangle}\mathbb{X}$.

The transition probabilities of the reduced semi-Markov process ${}_{\langle 1,N\rangle}\eta^{(\varepsilon)}(t)$ have, for every $\varepsilon \in (0, \varepsilon_0]$, the following form, for $t \geq 0$,

$${}_{\langle 1,N\rangle}Q_{ij}^{(\varepsilon)}(t) = \begin{cases} F_{1,+}^{(\varepsilon)}(t)\, p_{1,+}(\varepsilon) & \text{if } j = 2, i = 1, \\ {}_{\langle 1,N\rangle}F_{1,-}^{(\varepsilon)}(t)\, p_{1,-}(\varepsilon) & \text{if } j = 1, i = 1, \\ F_{i,\pm}^{(\varepsilon)}(t)\, p_{i,\pm}(\varepsilon) & \text{if } j = i \pm 1, 1 < i < N, \\ F_{N,\pm}^{(\varepsilon)}(t)\, p_{N,\pm}(\varepsilon) & \text{if } j = N - \frac{1\mp 1}{2}, i = N, \\ 0 & \text{otherwise,} \end{cases} \tag{11.75}$$

where

$${}_{\langle 1,N\rangle}F_{1,-}^{(\varepsilon)}(t) = \sum_{n=0}^{\infty} F_{1,-}^{(\varepsilon)}(t) * F_{0,-}^{(\varepsilon)*n}(t) * F_{0,+}^{(\varepsilon)}(t) \cdot p_{0,-}(\varepsilon)^n p_{0,+}(\varepsilon). \tag{11.76}$$

This relation implies, for every $\varepsilon \in (0, \varepsilon_0]$, the following relation for transition probabilities of the reduced embedded Markov chain ${}_{\langle 1,N\rangle}\eta_n^{(\varepsilon)}$,

$$
{}_{\langle 1,N\rangle}p_{ij}^{(\varepsilon)} = \begin{cases} {}_{\langle 1,N\rangle}p_{1,\pm}(\varepsilon) = p_{1,\pm}(\varepsilon) & \text{if } j = 1 + \frac{1\pm 1}{2}, i = 1, \\ {}_{\langle 1,N\rangle}p_{i,+}(\varepsilon) = p_{i,\pm}(\varepsilon) & \text{if } j = i \pm 1, 1 < i < N, \\ {}_{\langle 1,N\rangle}p_{N,+}(\varepsilon) = p_{N,\pm}(\varepsilon) & \text{if } j = N - \frac{1\mp 1}{2}, i = N, \\ 0 & \text{otherwise,} \end{cases} \tag{11.77}
$$

and the following relation for transition expectations of the reduced embedded semi-Markov process ${}_{\langle 1,N\rangle}\eta^{(\varepsilon)}(t)$,

$$
{}_{\langle 1,N\rangle}e_{ij}^{(\varepsilon)} = \begin{cases} {}_{\langle 1,N\rangle}e_{1,+}(\varepsilon) = e_{1,+}(\varepsilon) & \text{if } j = 2, i = 1, \\ {}_{\langle 1,N\rangle}e_{1,-}(\varepsilon) = e_{1,-}(\varepsilon) & \\ \quad + e_0(\varepsilon) \cdot \frac{p_{1,-}(\varepsilon)}{p_{0,+}(\varepsilon)} & \text{if } j = 1, i = 1, \\ {}_{\langle 1,N\rangle}e_{i,\pm}(\varepsilon) = e_{i,\pm}(\varepsilon) & \text{if } j = i \pm 1, 1 < i < N, \\ {}_{\langle 1,N\rangle}e_{N,\pm}(\varepsilon) = e_{N,\pm}(\varepsilon) & \text{if } j = N - \frac{1\mp 1}{2}, i = N, \\ 0 & \text{otherwise.} \end{cases} \tag{11.78}
$$

Note that, by Theorem 11.1, the following relation takes place, for $i, j \in {}_{\langle 1,N\rangle}\mathbb{X}$ and every $\varepsilon \in (0, \varepsilon_0]$,

$$
\mathsf{E}_i \tau_j^{(\varepsilon)} = \mathsf{E}_i \, {}_{\langle 1,N\rangle}\tau_j^{(\varepsilon)}. \tag{11.79}
$$

Analogously, the right end state N can be excluded from the phase space $\mathbb{X}$. In this case, the reduced phase space ${}_{\langle 0,N-1\rangle}\mathbb{X} = \{0, \ldots, N-1\}$.

As was mentioned above, the reduced semi-Markov processes ${}_{\langle 1,N\rangle}\eta^{(\varepsilon)}(t)$ and ${}_{\langle 0,N-1\rangle}\eta^{(\varepsilon)}(t)$ also have a birth-death type.

Let $0 \le k \le i \le r \le N$. The states $0, \ldots, k-1$ and $N, \ldots, r+1$ can be sequentially excluded from the phase space $\mathbb{X}$ of the semi-Markov process $\eta^{(\varepsilon)}(t)$.

Let us describe the corresponding recurrent procedure.

The reduced semi-Markov process ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$ can be obtained by excluding the state $k-1$ from the phase space ${}_{\langle k-1,j\rangle}\mathbb{X}$ of the reduced semi-Markov process ${}_{\langle k-1,r\rangle}\eta^{(\varepsilon)}(t)$ or by excluding state $r+1$ from the phase space ${}_{\langle k,r+1\rangle}\mathbb{X}$ of the reduced semi-Markov process ${}_{\langle k,r+1\rangle}\eta^{(\varepsilon)}(t)$.

The sequential exclusion of the states $0, \ldots, k-1$ and $N, \ldots, r+1$ can be realized in an arbitrary order of choice of one of these sequences and then by excluding the corresponding next state from the chosen sequence.

The simplest variants for the sequences of excluded states are $0, \ldots, k-1, N, \ldots, r+1$ and $N, \ldots, r+1, 0, \ldots, k-1$.

The resulting reduced semi-Markov process ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$ will be the same and it will have a birth-death type.

Here, we also accept the reduced semi-Markov process ${}_{\langle i,i\rangle}\eta^{(\varepsilon)}(t)$ with one-state phase space ${}_{\langle i,i\rangle}\mathbb{X} = \{i\}$ as a semi-Markov birth-death process.

This process has transition probability for the embedded Markov chain,

$$ {}_{\langle i,i\rangle}p_{ii}^{(\varepsilon)} = {}_{\langle i,i\rangle}p_{i,+}(\varepsilon) + {}_{\langle i,i\rangle}p_{i,-}(\varepsilon) = 1, \tag{11.80} $$

and the semi-Markov transition probabilities,

$$ \begin{aligned} {}_{\langle i,i\rangle}Q_{ii}^{(\varepsilon)}(t) &= {}_{\langle i,i\rangle}F_{i,+}^{(\varepsilon)}(t)\,{}_{\langle i,i\rangle}p_{i,+}(\varepsilon) + {}_{\langle i,i\rangle}F_{i,-}^{(\varepsilon)}(t)\,{}_{\langle i,i\rangle}p_{i,-}(\varepsilon). \\ &= \mathsf{P}_i\{\tau_i^{(\varepsilon)} \le t\}. \end{aligned} \tag{11.81} $$

The following relations, which are, in fact, variants of relations (11.77) and (11.78), express the transition probabilities ${}_{\langle k,r\rangle}p_{ij}^{(\varepsilon)}$ and the expectations of transition times ${}_{\langle k,r\rangle}e_{ij}^{(\varepsilon)}$ for the reduced semi-Markov process ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$, via the transition probabilities ${}_{\langle k-1,r\rangle}p_{ij}^{(\varepsilon)}$ and the expectations of transition times ${}_{\langle k-1,r\rangle}e_{ij}^{(\varepsilon)}$ for the reduced semi-Markov process ${}_{\langle k-1,r\rangle}\eta^{(\varepsilon)}(t)$, for $1 \le k \le r \le N$ and, for every $\varepsilon \in (0, \varepsilon_0]$,

$$ {}_{\langle k,r\rangle}p_{ij}^{(\varepsilon)} = \begin{cases} {}_{\langle k,r\rangle}p_{k,\pm}(\varepsilon) = {}_{\langle k-1,r\rangle}p_{k,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = k + \frac{1\pm1}{2}, i = k, \\ {}_{\langle k,r\rangle}p_{i,\pm}(\varepsilon) = {}_{\langle k-1,r\rangle}p_{i,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = i \pm 1, k < i < r, \\ {}_{\langle k,r\rangle}p_{r,\pm}(\varepsilon) = {}_{\langle k-1,r\rangle}p_{r,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = r - \frac{1\mp1}{2}, i = r, \\ 0 \qquad\qquad \text{otherwise}, \end{cases} \tag{11.82} $$

and

$$ {}_{\langle k,r\rangle}e_{ij}^{(\varepsilon)} = \begin{cases} {}_{\langle k,r\rangle}e_{k,+}(\varepsilon) = {}_{\langle k-1,r\rangle}e_{k,+}(\varepsilon) \\ \qquad\qquad \text{if } j = k+1, i = k, \\ {}_{\langle k,r\rangle}e_{k,-}(\varepsilon) = {}_{\langle k-1,r\rangle}e_{k,-}(\varepsilon) \\ \qquad\qquad + {}_{\langle k-1,r\rangle}e_{k-1}(\varepsilon) \cdot \frac{{}_{\langle k-1,r\rangle}p_{k,-}(\varepsilon)}{{}_{\langle k-1,r\rangle}p_{k-1,+}(\varepsilon)} \\ \qquad\qquad \text{if } j = k, i = k, \\ {}_{\langle k,r\rangle}e_{i,\pm}(\varepsilon) = {}_{\langle k-1,r\rangle}e_{i,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = i \pm 1, k < i < r, \\ {}_{\langle k,r\rangle}e_{r,\pm}(\varepsilon) = {}_{\langle k-1,r\rangle}e_{r,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = r - \frac{1\mp1}{2}, i = r, \\ 0 \qquad\qquad \text{otherwise}, \end{cases} \tag{11.83} $$

where

$$ {}_{\langle k,r\rangle}e_i(\varepsilon) = {}_{\langle k,r\rangle}e_{i,+}(\varepsilon) + {}_{\langle k,r\rangle}e_{i,-}(\varepsilon). \tag{11.84} $$

The transition probabilities ${}_{\langle k,r\rangle}p_{ij}^{(\varepsilon)}$ and the expectations of transition times ${}_{\langle k,r\rangle}e_{ij}^{(\varepsilon)}$ for the reduced semi-Markov process ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$ can also be expressed via the transition probabilities ${}_{\langle k,r+1\rangle}p_{ij}^{(\varepsilon)}$ and the expectations of transition times ${}_{\langle k,r+1\rangle}e_{ij}^{(\varepsilon)}$ for the reduced semi-Markov process ${}_{\langle k,r+1\rangle}\eta^{(\varepsilon)}(t)$, for $0 \le k \le r \le N-1$ in an analogous way.

11.5.4 Explicit Formulas for Expectations of Hitting Times for Semi-Markov Birth-Death Processes

By iterating the recurrent formulas (11.82)–(11.83) and their right hand analogues, we get the following explicit formulas for the transition probabilities ${}_{\langle k,r\rangle}p_{ij}^{(\varepsilon)}$ and the expectations of transition times ${}_{\langle k,r\rangle}e_{ij}^{(\varepsilon)}$ for the reduced semi-Markov process ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$ expressed in terms of the transition characteristics for the initial semi-Markov process $\eta^{(\varepsilon)}(t)$, for $0 \le k \le r \le N$ and, for every $\varepsilon \in (0, \varepsilon_0]$,

$$
{}_{\langle k,r\rangle}p_{ij}^{(\varepsilon)} = \begin{cases} {}_{\langle k,r\rangle}p_{k,\pm}(\varepsilon) = p_{k,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = k + \frac{1\pm 1}{2}, i = k, \\ {}_{\langle k,r\rangle}p_{i,\pm}(\varepsilon) = p_{i,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = i \pm 1, k < i < r, \\ {}_{\langle k,r\rangle}p_{r,+}(\varepsilon) = p_{r,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = r - \frac{1\mp 1}{2}, i = r, \\ 0 \qquad\qquad \text{otherwise.} \end{cases} \tag{11.85}
$$

and

$$
{}_{\langle k,r\rangle}e_{ij}^{(\varepsilon)} = \begin{cases} {}_{\langle k,r\rangle}e_{k,+}(\varepsilon) = e_{k,+}(\varepsilon) \\ \qquad\qquad \text{if } j = k+1, i = k, \\ {}_{\langle k,r\rangle}e_{k,-}(\varepsilon) = e_{k,-}(\varepsilon) + e_{k-1}(\varepsilon) \cdot \frac{p_{k,-}(\varepsilon)}{p_{k-1,+}(\varepsilon)} \\ \qquad\qquad + \cdots + e_0(\varepsilon) \cdot \frac{p_{1,-}(\varepsilon)\cdots p_{k,-}(\varepsilon)}{p_{0,+}(\varepsilon)\cdots p_{k-1,+}(\varepsilon)} \\ \qquad\qquad \text{if } j = k, i = k, \\ {}_{\langle k,r\rangle}e_{i,+}(\varepsilon) = e_{i,\pm}(\varepsilon) \\ \qquad\qquad \text{if } j = i \pm 1, k < i < r, \\ {}_{\langle k,r\rangle}e_{r,+}(\varepsilon) = e_{r,+}(\varepsilon) + e_{r+1}(\varepsilon) \cdot \frac{p_{r,+}(\varepsilon)}{p_{r+1,-}(\varepsilon)} \\ \qquad\qquad + \cdots + e_N(\varepsilon) \cdot \frac{p_{N-1,+}(\varepsilon)\cdots p_{r,+}(\varepsilon)}{p_{N,-}(\varepsilon)\cdots p_{r+1,-}(\varepsilon)} \\ \qquad\qquad \text{if } j = r, i = r, \\ {}_{\langle k,r\rangle}e_{r,-}(\varepsilon) = e_{r,-}(\varepsilon) \\ \qquad\qquad \text{if } j = r-1, i = r, \\ 0 \qquad\qquad \text{otherwise.} \end{cases} \tag{11.86}
$$

Recall that ${}_{\langle k,r\rangle}\tau_j^{(\varepsilon)}$ is the hitting time for the state $j \in {}_{\langle k,r\rangle}\mathbb{X}$ for the reduced semi-Markov process ${}_{\langle k,r\rangle}\eta^{(\varepsilon)}(t)$.

By Theorem 11.1, the following relation takes place, for $i, j \in {}_{\langle k,r\rangle}\mathbb{X}$ and, for every $\varepsilon \in (0, \varepsilon_0]$,

$$
\mathsf{E}_i \tau_j^{(\varepsilon)} = \mathsf{E}_i \, {}_{\langle k,r\rangle}\tau_j^{(\varepsilon)}. \tag{11.87}
$$

Let us now choose $k = r = i \in \mathbb{X}$. In this case, the reduced phase space ${}_{\langle i,i\rangle}\mathbb{X} = \{i\}$ is a one-state set. In this case, the process ${}_{\langle i,i\rangle}\eta^{(\varepsilon)}(t)$ returns to the state i after every jump. This implies that, in this case, for every $\varepsilon \in (0, \varepsilon_0]$,

$$E_{ii}(\varepsilon) = \mathsf{E}_i \tau_i^{(\varepsilon)} = \mathsf{E}_{i\,\langle i,i\rangle} \tau_i^{(\varepsilon)} = {}_{\langle i,i\rangle} e_i(\varepsilon). \tag{11.88}$$

Thus, the following formulas take place, for every $i \in \mathbb{X}$ and, for every $\varepsilon \in (0, \varepsilon_0]$,

$$\begin{aligned} E_{ii}(\varepsilon) &= e_i(\varepsilon) \\ &+ e_{i-1}(\varepsilon)\frac{p_{i,-}(\varepsilon)}{p_{i-1,+}(\varepsilon)} + e_{i-2}(\varepsilon)\frac{p_{i-1,-}(\varepsilon)p_{i,-}(\varepsilon)}{p_{i-2,+}(\varepsilon)p_{i-1,+}(\varepsilon)} \\ &+ \cdots + e_0(\varepsilon)\frac{p_{1,-}(\varepsilon)p_{2,-}(\varepsilon)\cdots p_{i,-}(\varepsilon)}{p_{0,+}(\varepsilon)p_{1,+}(\varepsilon)\cdots p_{i-1,+}(\varepsilon)} \\ &+ e_{i+1}(\varepsilon)\frac{p_{i,+}(\varepsilon)}{p_{i+1,-}(\varepsilon)} + e_{i+2}(\varepsilon)\frac{p_{i+1,+}(\varepsilon)p_{i,+}(\varepsilon)}{p_{i+2,-}(\varepsilon)p_{i+1,-}(\varepsilon)} \\ &+ \cdots + e_N(\varepsilon)\frac{p_{N-1,+}(\varepsilon)p_{N-2,+}(\varepsilon)\cdots p_{i,+}(\varepsilon)}{p_{N,-}(\varepsilon)p_{N-1,-}(\varepsilon)\cdots p_{i+1,-}(\varepsilon)}. \end{aligned} \tag{11.89}$$

In what follows, we use the following well known formula for the stationary probabilities $\pi_i(\varepsilon)$, $i \in \mathbb{X}$, which takes place, for every $\varepsilon \in (0, \varepsilon_0]$,

$$\pi_i(\varepsilon) = \frac{e_i(\varepsilon)}{E_{ii}(\varepsilon)},\ i \in \mathbb{X}. \tag{11.90}$$

It should be noted that such formulas for stationary distributions of Markov birth-death chains are well known and can be found, for example, in Feller [11]. In context of our studies, a special value has the presented above recurrent algorithm for getting such formulas, based on sequential reduction of the phase space for semi-Markov birth-death processes.

As far as explicit expressions for conditional quasi-stationary probabilities are concerned, they can be obtained by substituting stationary probabilities $\pi_i(\varepsilon)$, $i \in \mathbb{X}$ given by formula (11.90) into formulas (11.37) and (11.38).

11.6 First and Second Order Asymptotic Expansions

In this section, we give explicit the first and the second order asymptotic expansions for stationary and conditional quasi-stationary distributions for perturbed semi-Markov birth-death processes.

The results of the present section are based on the explicit formula (11.89) for expected return times and the expressions which connect these quantities with stationary and conditional quasi-stationary distributions given respectively in formulas (11.90) and (11.37)–(11.38). We obtain the first and second order asymptotic expansions from these formulas by using operational rules for Laurent asymptotic expansions presented in Lemmas 11.1 and 11.2 given below.

It will be convenient to use the following notation,

$$\Gamma_{i,j,\pm}(\varepsilon) = p_{i,\pm}(\varepsilon)p_{i+1,\pm}(\varepsilon)\cdots p_{j,\pm}(\varepsilon),\ 0 \le i \le j \le N. \tag{11.91}$$

Using (11.91), we can write formula (11.89) as

$$E_{ii}(\varepsilon) = e_i(\varepsilon) + \sum_{k=0}^{i-1} e_k(\varepsilon)\frac{\Gamma_{k+1,i,-}(\varepsilon)}{\Gamma_{k,i-1,+}(\varepsilon)} + \sum_{k=i+1}^{N} e_k(\varepsilon)\frac{\Gamma_{i,k-1,+}(\varepsilon)}{\Gamma_{i+1,k,-}(\varepsilon)},\ i \in \mathbb{X}. \tag{11.92}$$

In particular, we have

$$E_{00}(\varepsilon) = e_0(\varepsilon) + \sum_{k \in {}_0\mathbb{X}} e_k(\varepsilon)\frac{\Gamma_{0,k-1,+}(\varepsilon)}{\Gamma_{1,k,-}(\varepsilon)}, \tag{11.93}$$

and

$$E_{NN}(\varepsilon) = e_N(\varepsilon) + \sum_{k \in {}_N\mathbb{X}} e_k(\varepsilon)\frac{\Gamma_{k+1,N,-}(\varepsilon)}{\Gamma_{k,N-1,+}(\varepsilon)}. \tag{11.94}$$

We will compute the desired asymptotic expansions by applying operational rules for Laurent asymptotic expansions in relations (11.92)–(11.94). In order for the presentation to not be too repetitive, we will directly compute the second order asymptotic expansions which contain the first order asymptotic expansions as special cases. In particular, this gives us limits for stationary and conditional quasi-stationary distributions.

The formulas for computing the asymptotic expansions are different depending on whether condition $\mathbf{H}_1$, $\mathbf{H}_2$ or $\mathbf{H}_3$ holds. We consider these three cases in Sects. 11.6.2, 11.6.3 and 11.6.4, respectively.

Each of these sections will have the same structure: First, we present a ma which successively constructs asymptotic expansions for the quantities given in relations (11.92)–(11.94). Then, using these expansions, we construct the first and the second order asymptotic expansions for stationary (Sects. 11.6.2–11.6.4) and conditional quasi-stationary distributions (Sects. 11.6.3–11.6.4).

11.6.1 Laurent Asymptotic Expansions

In this subsection, we present some operational rules for Laurent asymptotic expansions given in Silvestrov, D. and Silvestrov, S. [39–41] which are used in the present paper for constructions of asymptotic expansions for stationary and conditional quasi-stationary distributions of perturbed semi-Markov birth-death processes.

A real-valued function $A(\varepsilon)$, defined on an interval $(0, \varepsilon_0]$ for some $0 < \varepsilon_0 \le 1$, is a Laurent asymptotic expansion if it can be represented in the following form, $A(\varepsilon) = a_{h_A}\varepsilon^{h_A} + \cdots + a_{k_A}\varepsilon^{k_A} + o_A(\varepsilon^{k_A})$, $\varepsilon \in (0, \varepsilon_0]$, where **(a)** $-\infty < h_A \le k_A <$

∞ are integers, **(b)** coefficients $a_{h_A}, \dots, a_{k_A}$ are real numbers, **(c)** the function $o_A(\varepsilon^{k_A})/\varepsilon^{k_A} \to 0$ as $\varepsilon \to 0$. Such an expansion is pivotal if it is known that $a_{h_A} \neq 0$.

The above paper presents operational rules for Laurent asymptotic expansions. Let us shortly formulate some of these rules, in particular, for summation, multiplication and division of Laurent asymptotic expansions.

Lemma 11.1 *Let* $A(\varepsilon) = a_{h_A}\varepsilon^{h_A} + \cdots + a_{k_A}\varepsilon^{k_A} + o_A(\varepsilon^{k_A})$ *and* $B(\varepsilon) = b_{h_B}\varepsilon^{h_B} + \cdots + b_{k_B}\varepsilon^{k_B} + o_B(\varepsilon^{k_B})$ *be two pivotal Laurent asymptotic expansions. Then*:

(i) $C(\varepsilon) = cA(\varepsilon) = c_{h_C}\varepsilon^{h_C} + \cdots + c_{k_C}\varepsilon^{k_C} + o_C(\varepsilon^{k_C})$, *where a constant* $c \neq 0$, *is a pivotal Laurent asymptotic expansion and* $h_C = h_A, k_C = k_A, c_{h_C+r} = ca_{h_C+r}, r = 0, \dots, k_C - h_C$,

(ii) $D(\varepsilon) = A(\varepsilon) + B(\varepsilon) = d_{h_D}\varepsilon^{h_D} + \cdots + d_{k_D}\varepsilon^{k_D} + o_D(\varepsilon^{k_D})$ *is a pivotal Laurent asymptotic expansion and* $h_D = h_A \wedge h_B, k_D = k_A \wedge k_B, d_{h_D+r} = a_{h_D+r} + b_{h_D+r}, r = 0, \dots, k_D - h_D$, *where* $a_{h_D+r} = 0, r < h_A - h_D, b_{h_D+r} = 0, r < h_B - h_D$,

(iii) $E(\varepsilon) = A(\varepsilon) \cdot B(\varepsilon) = e_{h_E}\varepsilon^{h_E} + \cdots + e_{k_E}\varepsilon^{k_E} + o_E(\varepsilon^{k_E})$ *is a pivotal Laurent asymptotic expansion and* $h_E = h_A + h_B, k_E = (k_A + h_B) \wedge (k_B + h_A), e_{h_E+r} = \sum_{l=0}^{r} a_{h_A+l} \cdot b_{h_B+r-l}, r = 0, \dots, k_E - h_E$,

(iv) $F(\varepsilon) = A(\varepsilon)/B(\varepsilon) = f_{h_F}\varepsilon^{h_F} + \cdots + f_{k_F}\varepsilon^{k_F} + o_F(\varepsilon^{k_F})$ *ia a pivotal Laurent asymptotic expansion and* $h_F = h_A - h_B, k_F = (k_A - h_B) \wedge (k_B - 2h_B + h_A), f_{h_F+r} = \frac{1}{b_{h_B}}(a_{h_A+r} + \sum_{l=1}^{r} b_{h_B+l} \cdot f_{h_F+r-l}), r = 0, \dots, k_F - h_F$.

The following lemma presents useful multiple generalizations of summation and multiplication rules given in Lemma 11.1.

Lemma 11.2 *Let* $A_i(\varepsilon) = a_{i,h_{A_i}}\varepsilon^{h_A} + \cdots + a_{i,k_{A_i}}\varepsilon^{k_{A_i}} + o_{A_i}(\varepsilon^{k_{A_i}}), i = 1, \dots, m$ *be pivotal Laurent asymptotic expansions. Then*:

(i) $D(\varepsilon) = \sum_{i=1}^{m} A_i(\varepsilon) = d_{h_D}\varepsilon^{h_D} + \cdots + d_{k_D}\varepsilon^{k_D} + o_D(\varepsilon^{k_D})$ *is a pivotal Laurent asymptotic expansion and* $h_D = \min_{1 \le l \le m} h_{A_l}$, $k_D = \min_{1 \le l \le m} k_{A_l}$, $d_{h_D+l} = a_{1,h_D+l} + \cdots + a_{m,h_D+l}$, $l = 0, \dots, k_D - h_D$, *where* $a_{i,h_D+l} = 0$ *for* $0 \le l < h_{A_i} - h_D, i = 1, \dots, m$,

(ii) $E(\varepsilon) = \prod_{i=1}^{m} A_i(\varepsilon) = e_{h_E}\varepsilon^{h_E} + \cdots + e_{k_E}\varepsilon^{k_E} + o_E(\varepsilon^{k_E})$ *is a pivotal Laurent asymptotic expansion and*
$h_E = \sum_{l=1}^{m} h_{A_l}$, $k_E = \min_{1 \le l \le m}(k_{A_l} + \sum_{1 \le r \le m, r \neq l} h_{A_r})$,
$e_{h_E+l} = \sum_{l_1+\cdots+l_m=l, 0 \le l_i \le k_{A_i} - h_{A_i}, i=1,\dots,m} \prod_{1 \le i \le m} a_{i,h_{A_i}+l_i}$,
$l = 0, \dots, k_E - h_E$.

11.6.2 First and Second Order Asymptotic Expansions for Stationary Distributions Under Condition $\mathbf{H}_1$

In the case where condition $\mathbf{H}_1$ holds, the semi-Markov process has no asymptotically absorbing states. In this case, all quantities in relations (11.92)–(11.94) are of order $O(1)$ and the construction of asymptotic expansions are rather straightforward.

In the following lemma we successively construct asymptotic expansions for the quantities given in relations (11.92)–(11.94).

Lemma 11.3 *Assume that conditions* **A**–**G** *and* $\mathbf{H}_1$ *hold. Then:*

(i) *For* $i \in \mathbb{X}$, *we have,*

$$e_i(\varepsilon) = b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon),\ \varepsilon \in (0, \varepsilon_0],$$

where $\dot{o}_i(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$b_i[0] = b_{i,-}[0] + b_{i,+}[0] > 0, \quad b_i[1] = b_{i,-}[1] + b_{i,+}[1].$$

(ii) *For* $0 \le i \le j \le N$, *we have,*

$$\Gamma_{i,j,\pm}(\varepsilon) = A_{i,j,\pm}[0] + A_{i,j,\pm}[1]\varepsilon + o_{i,j,\pm}(\varepsilon),\ \varepsilon \in (0, \varepsilon_0],$$

where $o_{i,j,\pm}(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$A_{i,j,\pm}[0] = a_{i,\pm}[0]a_{i+1,\pm}[0]\cdots a_{j,\pm}[0] > 0,$$

$$A_{i,j,\pm}[1] = \sum_{n_i+n_{i+1}+\cdots+n_j=1} a_{i,\pm}[n_i]a_{i+1,\pm}[n_{i+1}]\cdots a_{j,\pm}[n_j].$$

(iii) *For* $0 \le k \le i-1,\ i \in {}_{\langle 1,N\rangle}\mathbb{X}$, *we have*

$$\frac{\Gamma_{k+1,i,-}(\varepsilon)}{\Gamma_{k,i-1,+}(\varepsilon)} = A^*_{k,i}[0] + A^*_{k,i}[1]\varepsilon + o^*_{k,i}(\varepsilon),\ \varepsilon \in (0, \varepsilon_0],$$

where $o^*_{k,i}(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$A^*_{k,i}[0] = \frac{A_{k+1,i,-}[0]}{A_{k,i-1,+}[0]} > 0,$$

$$A^*_{k,i}[1] = \frac{A_{k+1,i,-}[1]A_{k,i-1,+}[0] - A_{k+1,i,-}[0]A_{k,i-1,+}[1]}{A_{k,i-1,+}[0]^2}.$$

(iv) *For* $i+1 \le k \le N,\ i \in {}_{\langle 0,N-1\rangle}\mathbb{X}$, *we have*

$$\frac{\Gamma_{i,k-1,+}(\varepsilon)}{\Gamma_{i+1,k,-}(\varepsilon)} = A^*_{k,i}[0] + A^*_{k,i}[1]\varepsilon + o^*_{k,i}(\varepsilon),\ \varepsilon \in (0, \varepsilon_0],$$

where $o^*_{k,i}(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$A^*_{k,i}[0] = \frac{A_{i,k-1,+}[0]}{A_{i+1,k,-}[0]} > 0,$$

$$A^*_{k,i}[1] = \frac{A_{i,k-1,+}[1]A_{i+1,k,-}[0] - A_{i,k-1,+}[0]A_{i+1,k,-}[1]}{A_{i+1,k,-}[0]^2}.$$

(**v**) *For $i \in \mathbb{X}$, we have*

$$E_{ii}(\varepsilon) = B_{ii}[0] + B_{ii}[1]\varepsilon + \dot{o}_{ii}(\varepsilon),\ \varepsilon \in (0, \varepsilon_0],$$

where $\dot{o}_{ii}(\varepsilon)/\varepsilon \to 0$ as $\varepsilon \to 0$ and

$$B_{ii}[0] = b_i[0] + \sum_{k \in {}_i\mathbb{X}} b_k[0]A^*_{k,i}[0] > 0,$$

$$B_{ii}[1] = b_i[1] + \sum_{k \in {}_i\mathbb{X}} (b_k[0]A^*_{k,i}[1] + b_k[1]A^*_{k,i}[0]).$$

Proof Since $e_i(\varepsilon) = e_{i,-}(\varepsilon) + e_{i,+}(\varepsilon), i \in \mathbb{X}$, part (**i**) follows immediately from condition **E**.

For the proof of part (**ii**) we notice that it follows from the definition (11.91) of $\Gamma_{i,j,\pm}(\varepsilon)$ and condition **D** that

$$\Gamma_{i,j,\pm}(\varepsilon) = \prod_{k=i}^{j} (a_{k,\pm}[0] + a_{k,\pm}[1]\varepsilon + o_{k,\pm}(\varepsilon)),\ 0 \le i \le j \le N.$$

By applying the multiple product rule for asymptotic expansions, we obtain the asymptotic relation given in part (**ii**) where the coefficients $A_{i,j,\pm}[0], 0 \le i \le j \le n$, are positive since condition $\mathbf{H}_1$ holds.

In order to prove parts (**iii**) and (**iv**) we use the result in part (**ii**). For $0 \le k \le i-1$, $i \in {}_0\mathbb{X}$, this gives us

$$\frac{\Gamma_{k+1,i,-}(\varepsilon)}{\Gamma_{k,i-1,+}(\varepsilon)} = \frac{A_{k+1,i,-}[0] + A_{k+1,i,-}[1]\varepsilon + o_{k+1,i,-}(\varepsilon)}{A_{k,i-1,+}[0] + A_{k,i-1,+}[1]\varepsilon + o_{k,i-1,+}(\varepsilon)}, \tag{11.95}$$

and, for $i+1 \le k \le N$, $i \in {}_N\mathbb{X}$, we get

$$\frac{\Gamma_{i,k-1,+}(\varepsilon)}{\Gamma_{i+1,k,-}(\varepsilon)} = \frac{A_{i,k-1,+}[0] + A_{i,k-1,+}[1]\varepsilon + o_{i,k-1,+}(\varepsilon)}{A_{i+1,k,-}[0] + A_{i+1,k,-}[1]\varepsilon + o_{i+1,k,-}(\varepsilon)}. \tag{11.96}$$

Using the division rule for asymptotic expansions in relations (11.95) and (11.96) we get the asymptotic expansions given in parts (**iii**) and (**iv**).

Finally, we can use relation (11.92) to prove part (**v**). This relation together with the results in parts (**i**), (**iii**) and (**iv**) yield

$$E_{ii}(\varepsilon) = b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon) + \sum_{k \in \mathbb{X}\setminus\{i\}} (b_k[0] + b_k[1]\varepsilon + \dot{o}_k(\varepsilon)) \times (A^*_{k,i}[0] + A^*_{k,i}[1]\varepsilon + o^*_{k,i}(\varepsilon)),\ i \in \mathbb{X}.$$

A combination of the product rule and the multiple summation rule for asymptotic expansions gives the asymptotic relation in part (**v**). □

The following theorem gives second order asymptotic expansions for stationary probabilities. In particular, this theorem shows that there exist limits for stationary probabilities,

$$\pi_i(0) = \lim_{\varepsilon \to 0} \pi_i(\varepsilon),\ i \in \mathbb{X},$$

where $\pi_i(0) > 0$, $i \in \mathbb{X}$.

Theorem 11.2 *Assume that conditions* **A**–**G** *and* $\mathbf{H}_1$ *hold. Then, we have the following asymptotic relation for the stationary probabilities* $\pi_i(\varepsilon)$, $i \in \mathbb{X}$,

$$\pi_i(\varepsilon) = c_i[0] + c_i[1]\varepsilon + o_i(\varepsilon),\ \varepsilon \in (0, \varepsilon_0],$$

where $o_i(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$c_i[0] = \frac{b_i[0]}{B_{ii}[0]} > 0, \quad c_i[1] = \frac{b_i[1]B_{ii}[0] - b_i[0]B_{ii}[1]}{B_{ii}[0]^2},$$

where $B_{ii}[0]$, $B_{ii}[1]$, $i \in \mathbb{X}$, *can be computed from the formulas given in Lemma 11.3.*

Proof It follows from condition **E** and part (**v**) of Lemma 11.3 that, for $i \in \mathbb{X}$,

$$\pi_i(\varepsilon) = \frac{e_i(\varepsilon)}{E_{ii}(\varepsilon)} = \frac{b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon)}{B_{ii}[0] + B_{ii}[1]\varepsilon + \dot{o}_{ii}(\varepsilon)}.$$

The result now follows from the division rule for asymptotic expansions **(iv)**, given in Lemma 11.1. □

11.6.3 First and Second Order Asymptotic Expansions for Stationary and Conditional Quasi-stationary Distributions Under Condition $\mathbf{H}_2$

In the case where condition $\mathbf{H}_2$ holds, the semi-Markov process has one asymptotically absorbing state, namely state 0. This means that $p_{0,+}(\varepsilon) \sim O(\varepsilon)$ and since this quantity is involved in relations (11.92)–(11.94), the pivotal properties of the

expansions are less obvious. Furthermore, since some terms now tends to infinity, we partly need to operate with Laurent asymptotic expansions.

In order to separate cases where $i = 0$ or $i \in {}_{\langle 1,N\rangle}\mathbb{X}$ we will use the indicator function $\gamma_i = I(i = 0)$, that is, $\gamma_0 = 1$ and $\gamma_i = 0$ for $i \in {}_{\langle 1,N\rangle}\mathbb{X}$.

The following lemma gives asymptotic expansions for quantities in relations (11.92)–(11.94).

Lemma 11.4 *Assume that conditions* **A**–**G** *and* $\mathbf{H}_2$ *hold. Then:*

(i) *For* $i \in \mathbb{X}$, *we have,* $e_i(\varepsilon) = b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon)$, $\varepsilon \in (0, \varepsilon_0]$, *where* $\dot{o}_i(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$b_i[0] = b_{i,-}[0] + b_{i,+}[0] > 0, \quad b_i[1] = b_{i,-}[1] + b_{i,+}[1].$$

(ii) *For* $0 \le i \le j \le N$, *we have,*

$$\Gamma_{i,j,+}(\varepsilon) = A_{i,j,+}[\gamma_i]\varepsilon^{\gamma_i} + A_{i,j,+}[\gamma_i + 1]\varepsilon^{\gamma_i+1} + o_{i,j,+}(\varepsilon^{\gamma_i+1}),\ \varepsilon \in (0, \varepsilon_0],$$

where $o_{i,j,+}(\varepsilon^{\gamma_i+1})/\varepsilon^{\gamma_i+1} \to 0$ *as* $\varepsilon \to 0$ *and*

$$A_{i,j,+}[\gamma_i] = a_{i,+}[\gamma_i]a_{i+1,+}[0]\cdots a_{j,+}[0] > 0,$$

$$A_{i,j,+}[\gamma_i + 1] = \sum_{n_i+n_{i+1}+\cdots+n_j=1} a_{i,+}[\gamma_i + n_i]a_{i+1,+}[n_{i+1}]\cdots a_{j,+}[n_j].$$

(iii) *For* $0 \le i \le j \le N$, *we have,*

$$\Gamma_{i,j,-}(\varepsilon) = A_{i,j,-}[0] + A_{i,j,-}[1]\varepsilon + o_{i,j,-}(\varepsilon),\ \varepsilon \in (0, \varepsilon_0],$$

where $o_{i,j,-}(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$A_{i,j,-}[0] = a_{i,-}[0]a_{i+1,-}[0]\cdots a_{j,-}[0] > 0,$$

$$A_{i,j,-}[1] = \sum_{n_i+n_{i+1}+\cdots+n_j=1} a_{i,-}[n_i]a_{i+1,-}[n_{i+1}]\cdots a_{j,-}[n_j].$$

(iv) *For* $0 \le k \le i-1$, $i \in {}_{\langle 1,N\rangle}\mathbb{X}$, *we have,*

$$\frac{\Gamma_{k+1,i,-}(\varepsilon)}{\Gamma_{k,i-1,+}(\varepsilon)} = A^*_{k,i}[-\gamma_k]\varepsilon^{-\gamma_k} + A^*_{k,i}[-\gamma_k + 1]\varepsilon^{-\gamma_k+1} + o^*_{k,i}(\varepsilon^{-\gamma_k+1}),\ \varepsilon \in (0, \varepsilon_0],$$

where $o^*_{k,i}(\varepsilon^{-\gamma_k+1})/\varepsilon^{-\gamma_k+1} \to 0$ *as* $\varepsilon \to 0$ *and*

$$A^*_{k,i}[-\gamma_k] = \frac{A_{k+1,i,-}[0]}{A_{k,i-1,+}[\gamma_k]} > 0,$$

$$A^*_{k,i}[-\gamma_k + 1] = \frac{A_{k+1,i,-}[1]A_{k,i-1,+}[\gamma_k] - A_{k+1,i,-}[0]A_{k,i-1,+}[\gamma_k + 1]}{A_{k,i-1,+}[\gamma_k]^2}.$$

(v) *For $i + 1 \le k \le N$, $i \in {}_{\langle 0,N-1\rangle}\mathbb{X}$, we have,*

$$\frac{\Gamma_{i,k-1,+}(\varepsilon)}{\Gamma_{i+1,k,-}(\varepsilon)} = A^*_{k,i}[\gamma_i]\varepsilon^{\gamma_i} + A^*_{k,i}[\gamma_i + 1]\varepsilon^{\gamma_i+1} + o^*_{k,i}(\varepsilon^{\gamma_i+1}),\ \varepsilon \in (0, \varepsilon_0],$$

where $o^*_{k,i}(\varepsilon^{\gamma_i+1})/\varepsilon^{\gamma_i+1} \to 0$ as $\varepsilon \to 0$ and

$$A^*_{k,i}[\gamma_i] = \frac{A_{i,k-1,+}[\gamma_i]}{A_{i+1,k,-}[0]} > 0,$$

$$A^*_{k,i}[\gamma_i + 1] = \frac{A_{i,k-1,+}[\gamma_i + 1]A_{i+1,k,-}[0] - A_{i,k-1,+}[\gamma_i]A_{i+1,k,-}[1]}{A_{i+1,k,-}[0]^2}.$$

(vi) *For $i \in \mathbb{X}$, we have,*

$$E_{ii}(\varepsilon) = B_{ii}[\gamma_i - 1]\varepsilon^{\gamma_i-1} + B_{ii}[\gamma_i]\varepsilon^{\gamma_i} + \dot{o}_{ii}(\varepsilon^{\gamma_i}),\ \varepsilon \in (0, \varepsilon_0],$$

where $\dot{o}_{ii}(\varepsilon^{\gamma_i})/\varepsilon^{\gamma_i} \to 0$ as $\varepsilon \to 0$ and

$$B_{00}[0] = b_0[0] > 0, \quad B_{00}[1] = b_0[1] + \sum_{k \in {}_0\mathbb{X}} b_k[0]A^*_{k,0}[1],$$

$$B_{ii}[-1] = b_0[0]A^*_{0,i}[-1] > 0,\ i \in {}_{\langle 1,N\rangle}\mathbb{X},$$

$$B_{ii}[0] = b_0[1]A^*_{0,i}[-1] + b_i[0] + \sum_{k \in {}_i\mathbb{X}} b_k[0]A^*_{k,i}[0],\ i \in {}_{\langle 1,N\rangle}\mathbb{X}.$$

Proof Let us first note that the quantities in parts (**i**) and (**iii**) do not depend on $p_{0,+}(\varepsilon)$, so the proofs for these parts are the same as the proofs for parts (**i**) and (**ii**) in Lemma 11.3, respectively.

We now prove part (**ii**). Notice that it follows from conditions **D** and $\mathbf{H}_2$ that $p_{i,+}(\varepsilon) = a_{i,+}[\gamma_i]\varepsilon^{\gamma_i} + a_{i,+}[\gamma_i + 1]\varepsilon^{\gamma_i+1} + o_{i,+}(\varepsilon^{\gamma_i+1})$, $i \in \mathbb{X}$. Using this and the definition (11.91) of $\Gamma_{i,j,+}(\varepsilon)$ gives

$$\begin{aligned}\Gamma_{i,j,+}(\varepsilon) &= (a_{i,+}[\gamma_i]\varepsilon^{\gamma_i} + a_{i,+}[\gamma_i+1]\varepsilon^{\gamma_i+1} + o_{i,+}(\varepsilon^{\gamma_i+1})) \\ &\quad \times (a_{i+1,+}[0] + a_{i+1,+}[1]\varepsilon + o_{i+1,+}(\varepsilon)) \\ &\quad \times \cdots \times \\ &\quad \times (a_{j,+}[0] + a_{j,+}[1]\varepsilon + o_{j,+}(\varepsilon)),\ 0 \le i \le j \le N.\end{aligned}$$

An application of the multiple product rule for asymptotic expansions shows that part (**ii**) holds.

Now, using the results in parts (**ii**) and (**iii**) we get, for $0 \le k \le i-1$, $i \in {}_{\langle 1,N\rangle}\mathbb{X}$,

$$\frac{\Gamma_{k+1,i,-}(\varepsilon)}{\Gamma_{k,i-1,+}(\varepsilon)} = \frac{A_{k+1,i,-}[0] + A_{k+1,i,-}[1]\varepsilon + o_{k+1,i,-}(\varepsilon)}{A_{k,i-1,+}[\gamma_k]\varepsilon^{\gamma_k} + A_{k,i-1,+}[\gamma_k+1]\varepsilon^{\gamma_k+1} + o_{k,i-1,+}(\varepsilon^{\gamma_k+1})},$$

and, for $i+1 \le k \le N$, $i \in {}_{N}\mathbb{X}$,

$$\frac{\Gamma_{i,k-1,+}(\varepsilon)}{\Gamma_{i+1,k,-}(\varepsilon)} = \frac{A_{i,k-1,+}[\gamma_i]\varepsilon^{\gamma_i} + A_{i,k-1,+}[\gamma_i+1]\varepsilon^{\gamma_i+1} + o_{i,k-1,+}(\varepsilon^{\gamma_i+1})}{A_{i+1,k,-}[0] + A_{i+1,k,-}[1]\varepsilon + o_{i+1,k,-}(\varepsilon)}.$$

Notice that it is possible that the quantity in the first of the above equations tends to infinity as $\varepsilon \to 0$. Applying the division rule for Laurent asymptotic expansions in the above two relations yields the asymptotic relations given in parts (**iv**) and (**v**).

In order to prove part (**vi**), we consider the cases $i = 0$ and $i \in {}_{\langle 1,N\rangle}\mathbb{X}$ separately. First note that it follows from relation (11.93) and the results in parts (**i**) and (**iv**) that

$$\begin{aligned}E_{00}(\varepsilon) &= b_0[0] + b_0[1]\varepsilon + \dot{o}_0(\varepsilon) \\ &\quad + \sum_{k \in {}_0\mathbb{X}} (b_k[0] + b_k[1]\varepsilon + \dot{o}_k(\varepsilon))(A^*_{k,0}[1]\varepsilon + A^*_{k,0}[2]\varepsilon^2 + o^*_{k,0}(\varepsilon^2)).\end{aligned}$$

Using the product rule and the multiple summation rule for asymptotic expansions we obtain the asymptotic relation in part (**vi**) for the case $i = 0$.

If $i \in {}_{\langle 1,N\rangle}\mathbb{X}$, relation (11.92) implies together with parts (**i**), (**iv**) and (**v**) that

$$\begin{aligned}E_{ii}(\varepsilon) &= b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon) \\ &\quad + (b_0[0] + b_0[1]\varepsilon + \dot{o}_0(\varepsilon))(A^*_{0,i}[-1]\varepsilon^{-1} + A^*_{0,i}[0] + o^*_{0,i}(1)) \\ &\quad + \sum_{k \in {}_{0,i}\mathbb{X}} (b_k[0] + b_k[1]\varepsilon + \dot{o}_k(\varepsilon))(A^*_{k,i}[0] + A^*_{k,i}[1]\varepsilon + o^*_{k,i}(\varepsilon)).\end{aligned}$$

Notice that the term corresponding to $k = 0$ is of order $O(\varepsilon^{-1})$ while all other terms in the sum are of order $O(1)$. We can again apply the product rule and multiple summation rule for Laurent asymptotic expansions and in this case, the asymptotic relation in part (**vi**) is obtained for $i \in {}_{\langle 1,N\rangle}\mathbb{X}$. □

The following theorem gives second order asymptotic expansions for stationary and conditional quasi-stationary probabilities. In particular, part (**i**) of this theorem

shows that there exist limits for stationary probabilities, $\pi_i(0) = \lim_{\varepsilon \to 0} \pi_i(\varepsilon), i \in \mathbb{X}$, where $\pi_0(0) = 1$ and $\pi_i(0) = 0$ for $i \in {}_{\langle 1,N \rangle}\mathbb{X}$.

Furthermore, part (**ii**) of the theorem shows, in particular, that there exist limits for conditional quasi-stationary probabilities $\tilde{\pi}_i(0) = \lim_{\varepsilon \to 0} \tilde{\pi}_i(\varepsilon)$, $i \in {}_{\langle 1,N \rangle}\mathbb{X}$, where $\tilde{\pi}_i(0) > 0, i \in {}_{\langle 1,N \rangle}\mathbb{X}$.

Theorem 11.3 *Assume that conditions* **A**–**G** *and* $\mathbf{H}_2$ *hold. Then:*

(**i**) *We have the following asymptotic relation for the stationary probabilities* $\pi_i(\varepsilon)$, $i \in \mathbb{X}$,

$$\pi_i(\varepsilon) = c_i[\tilde{l}_i]\varepsilon^{\tilde{l}_i} + c_i[\tilde{l}_i + 1]\varepsilon^{\tilde{l}_i+1} + o_i(\varepsilon^{\tilde{l}_i+1}),\ \varepsilon \in (0, \varepsilon_0],$$

where $\tilde{l}_i = I(i \neq 0)$, $o_i(\varepsilon^{\tilde{l}_i+1})/\varepsilon^{\tilde{l}_i+1} \to 0$ *as* $\varepsilon \to 0$, *and*

$$c_i[\tilde{l}_i] = \frac{b_i[0]}{B_{ii}[-\tilde{l}_i]} > 0, \quad c_i[\tilde{l}_i + 1] = \frac{b_i[1]B_{ii}[-\tilde{l}_i] - b_i[0]B_{ii}[-\tilde{l}_i + 1]}{B_{ii}[-\tilde{l}_i]^2},$$

where $B_{ii}[-1]$, $i \in {}_{\langle 1,N \rangle}\mathbb{X}$, $B_{ii}[0]$, $i \in \mathbb{X}$, *and* $B_{00}[1]$, *can be computed from the formulas given in Lemma 11.4.*

(**ii**) *We have the following asymptotic relation for the conditional quasi-stationary probabilities* $\tilde{\pi}_i(\varepsilon)$, $i \in {}_{\langle 1,N \rangle}\mathbb{X}$,

$$\tilde{\pi}_i(\varepsilon) = \tilde{c}_i[0] + \tilde{c}_i[1]\varepsilon + \tilde{o}_i(\varepsilon),\ \varepsilon \in (0, \varepsilon_0],$$

where $\tilde{o}_i(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$\tilde{c}_i[0] = \frac{c_i[1]}{d[1]} > 0, \quad \tilde{c}_i[1] = \frac{c_i[2]d[1] - c_i[1]d[2]}{d[1]^2},$$

where $d[l] = \sum_{j \in {}_0\mathbb{X}} c_i[l],\ l = 1, 2.$

Proof It follows from parts (**i**) and (**vi**) in Lemma 11.4 that, for $i \in \mathbb{X}$,

$$\pi_i(\varepsilon) = \frac{e_i(\varepsilon)}{E_{ii}(\varepsilon)} = \frac{b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon)}{B_{ii}[\gamma_i - 1]\varepsilon^{\gamma_i - 1} + B_{ii}[\gamma_i]\varepsilon^{\gamma_i} + \dot{o}_{ii}(\varepsilon^{\gamma_i})}. \tag{11.97}$$

We also have $\gamma_i = I(i = 0) = 1 - I(i \neq 0) = 1 - \tilde{l}_i$. By changing the indicator function and then using the division rule for Laurent asymptotic expansions in relation (11.97), we obtain the asymptotic expansion given in part (**i**).

In order to prove part (**ii**) we first use part (**i**) for $i \in {}_{\langle 1,N \rangle}\mathbb{X}$ to get

$$\tilde{\pi}_i(\varepsilon) = \frac{\pi_i(\varepsilon)}{\sum_{j \in {}_0\mathbb{X}} \pi_j(\varepsilon)} = \frac{c_i[1]\varepsilon + c_i[2]\varepsilon^2 + o_i(\varepsilon^2)}{\sum_{j \in {}_0\mathbb{X}}(c_j[1]\varepsilon + c_j[2]\varepsilon^2 + o_j(\varepsilon^2))},$$

and then we apply the multiple summation rule **(i)** for asymptotic expansions, given in Lemma 11.2, and the division rule for asymptotic expansions, given in Lemma 11.1. □

11.6.4 First and Second Order Asymptotic Expansions for Stationary and Conditional Quasi-stationary Distributions Under Condition $\mathbf{H_3}$

In the case where condition $\mathbf{H}_3$ holds, both state 0 and state N are asymptotically absorbing for the semi-Markov process. This means that $p_{0,+}(\varepsilon) \sim O(\varepsilon)$ and $p_{N,-}(\varepsilon) \sim O(\varepsilon)$ which makes the asymptotic analysis of relations (11.92)–(11.94) even more involved.

In order to separate cases where $i = 0$, $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$ or $i = N$, we will use the indicator functions $\gamma_i = I(i = 0)$ and $\beta_i = I(i = N)$.

The following lemma gives asymptotic expansions for quantities given in relations (11.92)–(11.94).

Lemma 11.5 *Assume that conditions* **A**–**G** *and* $\mathbf{H}_3$ *hold. Then:*

(i) *For* $i \in \mathbb{X}$, *we have* $e_i(\varepsilon) = b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon)$, $\varepsilon \in (0, \varepsilon_0]$, *where* $\dot{o}_i(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$b_i[0] = b_{i,-}[0] + b_{i,+}[0] > 0, \quad b_i[1] = b_{i,-}[1] + b_{i,+}[1].$$

(ii) *For* $0 \le i \le j \le N$, *we have*

$$\Gamma_{i,j,+}(\varepsilon) = A_{i,j,+}[\gamma_i]\varepsilon^{\gamma_i} + A_{i,j,+}[\gamma_i + 1]\varepsilon^{\gamma_i+1} + o_{i,j,+}(\varepsilon^{\gamma_i+1}), \ \varepsilon \in (0, \varepsilon_0],$$

where $o_{i,j,+}(\varepsilon^{\gamma_i+1})/\varepsilon^{\gamma_i+1} \to 0$ *as* $\varepsilon \to 0$ *and*

$$A_{i,j,+}[\gamma_i] = a_{i,+}[\gamma_i]a_{i+1,+}[0] \cdots a_{j,+}[0] > 0,$$

$$A_{i,j,+}[\gamma_i + 1] = \sum_{n_i+n_{i+1}+\cdots+n_j=1} a_{i,+}[\gamma_i + n_i]a_{i+1,+}[n_{i+1}] \cdots a_{j,+}[n_j].$$

(iii) *For* $0 \le i \le j \le N$, *we have*

$$\Gamma_{i,j,-}(\varepsilon) = A_{i,j,-}[\beta_j]\varepsilon^{\beta_j} + A_{i,j,-}[\beta_j + 1]\varepsilon^{\beta_j+1} + o_{i,j,-}(\varepsilon^{\beta_j+1}), \ \varepsilon \in (0, \varepsilon_0],$$

where $o_{i,j,-}(\varepsilon^{\beta_j+1})/\varepsilon^{\beta_j+1} \to 0$ *as* $\varepsilon \to 0$ *and*

$$A_{i,j,-}[\beta_j] = a_{i,-}[0] \cdots a_{j-1,-}[0]a_{j,-}[\beta_j] > 0,$$

$$A_{i,j,-}[\beta_j+1]=\sum_{n_i+\cdots+n_{j-1}+n_j=1} a_{i,-}[n_i]\cdots a_{j-1,-}[n_{j-1}]a_{j,-}[\beta_j+n_j].$$

(**iv**) *For* $0\le k\le i-1,\ i\in {}_{\langle 1,N\rangle}\mathbb{X}$, *we have*

$$\frac{\Gamma_{k+1,i,-}(\varepsilon)}{\Gamma_{k,i-1,+}(\varepsilon)}=A^*_{k,i}[\beta_i-\gamma_k]\varepsilon^{\beta_i-\gamma_k}+A^*_{k,i}[\beta_i-\gamma_k+1]\varepsilon^{\beta_i-\gamma_k+1}+o^*_{k,i}(\varepsilon^{\beta_i-\gamma_k+1}),\ \varepsilon\in(0,\varepsilon_0],$$

where $o^*_{k,i}(\varepsilon^{\beta_i-\gamma_k+1})/\varepsilon^{\beta_i-\gamma_k+1}\to 0$ *as* $\varepsilon\to 0$ *and*

$$A^*_{k,i}[\beta_i-\gamma_k]=\frac{A_{k+1,i,-}[\beta_i]}{A_{k,i-1,+}[\gamma_k]}>0,$$

$$A^*_{k,i}[\beta_i-\gamma_k+1]=\frac{A_{k+1,i,-}[\beta_i+1]A_{k,i-1,+}[\gamma_k]-A_{k+1,i,-}[\beta_i]A_{k,i-1,+}[\gamma_k+1]}{A_{k,i-1,+}[\gamma_k]^2}.$$

(**v**) *For* $i+1\le k\le N,\ i\in {}_{\langle 0,N-1\rangle}\mathbb{X}$, *we have*

$$\frac{\Gamma_{i,k-1,+}(\varepsilon)}{\Gamma_{i+1,k,-}(\varepsilon)}=A^*_{k,i}[\gamma_i-\beta_k]\varepsilon^{\gamma_i-\beta_k}+A^*_{k,i}[\gamma_i-\beta_k+1]\varepsilon^{\gamma_i-\beta_k+1}+o^*_{k,i}(\varepsilon^{\gamma_i-\beta_k+1}),\ \varepsilon\in(0,\varepsilon_0],$$

where $o^*_{k,i}(\varepsilon^{\gamma_i-\beta_k+1})/\varepsilon^{\gamma_i-\beta_k+1}\to 0$ *as* $\varepsilon\to 0$ *and*

$$A^*_{k,i}[\gamma_i-\beta_k]=\frac{A_{i,k-1,+}[\gamma_i]}{A_{i+1,k,-}[\beta_k]}>0,$$

$$A^*_{k,i}[\gamma_i-\beta_k+1]=\frac{A_{i,k-1,+}[\gamma_i+1]A_{i+1,k,-}[\beta_k]-A_{i,k-1,+}[\gamma_i]A_{i+1,k,-}[\beta_k+1]}{A_{i+1,k,-}[\beta_k]^2}.$$

(**vi**) *For* $i\in\mathbb{X}$, *we have*

$$E_{ii}(\varepsilon)=B_{ii}[\gamma_i+\beta_i-1]\varepsilon^{\gamma_i+\beta_i-1}+B_{ii}[\gamma_i+\beta_i]\varepsilon^{\gamma_i+\beta_i}+\dot o_{ii}(\varepsilon^{\gamma_i+\beta_i}),\ \varepsilon\in(0,\varepsilon_0],$$

where $\dot o_{ii}(\varepsilon^{\gamma_i+\beta_i})/\varepsilon^{\gamma_i+\beta_i}\to 0$ *as* $\varepsilon\to 0$ *and*

$$B_{ii}[0]=b_i[0]+b_{N-i}[0]A^*_{N-i,i}[0]>0,\ i=0,N,$$

$$B_{ii}[1] = b_{N-i}[1]A^*_{N-i,i}[0] + b_i[1] + \sum_{k \in {}_i\mathbb{X}} b_k[0]A^*_{k,i}[1],\ i = 0, N,$$

$$B_{ii}[-1] = b_0[0]A^*_{0,i}[-1] + b_N[0]A^*_{N,i}[-1] > 0,\ i \in {}_{\langle 1,N-1\rangle}\mathbb{X},$$

$$B_{ii}[0] = b_0[1]A^*_{0,i}[-1] + b_N[1]A^*_{N,i}[-1] + b_i[0]$$

$$+ \sum_{k \in {}_i\mathbb{X}} b_k[0]A^*_{k,i}[0],\ i \in {}_{\langle 1,N-1\rangle}\mathbb{X}.$$

Proof We first note that the quantities in parts (**i**) and (**ii**) do not depend on $p_{N,-}(\varepsilon)$, so the proofs for these parts are the same as the proofs for parts (**i**) and (**ii**) in Lemma 11.4, respectively.

In order to prove part (**iii**) we notice that it follows from conditions **D** and $\mathbf{H}_3$ that $p_{i,-}(\varepsilon) = a_{i,-}[\beta_i]\varepsilon^{\beta_i} + a_{i,-}[\beta_i + 1]\varepsilon^{\beta_i+1} + o_{i,-}(\varepsilon^{\beta_i+1})$, $i \in \mathbb{X}$. From this and the definition (11.91) of $\Gamma_{i,j,-}(\varepsilon)$ we get, for $0 \le i \le j \le N$,

$$\begin{aligned}\Gamma_{i,j,-}(\varepsilon) &= (a_{i,-}[0] + a_{i,-}[1]\varepsilon + o_{i,-}(\varepsilon)) \times \cdots \times \\ &\times (a_{j-1,-}[0] + a_{j-1,-}[1]\varepsilon + o_{j-1,-}(\varepsilon)) \\ &\times (a_{j,-}[\beta_j]\varepsilon^{\beta_j} + a_{j,-}[\beta_j + 1] + \varepsilon^{\beta_j+1} + o_{j,-}(\varepsilon^{\beta_j+1})).\end{aligned}$$

By applying the multiple product rule for asymptotic expansions we obtain the asymptotic relation given in part (**iii**).

From parts (**ii**) and (**iii**) it follows that, for $0 \le k \le i-1$, $i \in {}_{\langle 1,N\rangle}\mathbb{X}$,

$$\frac{\Gamma_{k+1,i,-}(\varepsilon)}{\Gamma_{k,i-1,+}(\varepsilon)} = \frac{A_{k+1,i,-}[\beta_i]\varepsilon^{\beta_i} + A_{k+1,i,-}[\beta_i + 1]\varepsilon^{\beta_i+1} + o_{k+1,i,-}(\varepsilon^{\beta_i+1})}{A_{k,i-1,+}[\gamma_k]\varepsilon^{\gamma_k} + A_{k,i-1,+}[\gamma_k + 1]\varepsilon^{\gamma_k+1} + o_{k,i-1,+}(\varepsilon^{\gamma_k+1})},$$

and, for $i + 1 \le k \le N$, $i \in {}_{\langle 0,N-1\rangle}\mathbb{X}$,

$$\frac{\Gamma_{i,k-1,+}(\varepsilon)}{\Gamma_{i+1,k,-}(\varepsilon)} = \frac{A_{i,k-1,+}[\gamma_i]\varepsilon^{\gamma_i} + A_{i,k-1,+}[\gamma_i + 1]\varepsilon^{\gamma_i+1} + o_{i,k-1,+}(\varepsilon^{\gamma_i+1})}{A_{i+1,k,-}[\beta_k]\varepsilon^{\beta_k} + A_{i+1,k,-}[\beta_k + 1]\varepsilon^{\beta_k+1} + o_{i+1,k,-}(\varepsilon^{\beta_k+1})}.$$

Notice that in the above two relations it is possible that the corresponding quantity tends to infinity as $\varepsilon \to 0$.

The asymptotic relations given in parts (**iv**) and (**v**) are obtained by using the division rule for Laurent asymptotic expansions in the above two relations.

We finally give the proof of part (**vi**). For the case $i = 0$, it follows from relation (11.93) and parts (**i**) and (**v**) that

$$E_{00}(\varepsilon) = b_0[0] + b_0[1]\varepsilon + \dot{o}_0(\varepsilon)$$
$$+ (b_N[0] + b_N[1]\varepsilon + \dot{o}_N(\varepsilon))(A^*_{N,0}[0] + A^*_{N,0}[1]\varepsilon + o^*_{N,0}(\varepsilon))$$
$$+ \sum_{k \in {}_{0,N}\mathbb{X}} (b_k[0] + b_k[1]\varepsilon + \dot{o}_k(\varepsilon))(A^*_{k,0}[1]\varepsilon + A^*_{k,0}[2]\varepsilon^2 + o^*_{k,0}(\varepsilon^2)).$$

The product rule and multiple summation rule for asymptotic expansions now proves part (**vi**) for the case $i = 0$.

If $i = N$, it follows from relation (11.94) and parts (**i**) and (**iv**) that

$$E_{NN}(\varepsilon) = b_N[0] + b_N[1]\varepsilon + \dot{o}_N(\varepsilon)$$
$$+ (b_0[0] + b_0[1]\varepsilon + \dot{o}_0(\varepsilon))(A^*_{0,N}[0] + A^*_{0,N}[1]\varepsilon + o^*_{0,N}(\varepsilon))$$
$$+ \sum_{k \in {}_{0,N}\mathbb{X}} (b_k[0] + b_k[1]\varepsilon + \dot{o}_k(\varepsilon))(A^*_{k,N}[1]\varepsilon + A^*_{k,N}[2]\varepsilon^2 + o^*_{k,N}(\varepsilon^2)).$$

Again, we can use the product rule and multiple summation rule in order to prove part (**vi**), in this case, for $i = N$.

For the case where $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$, we use relation (11.92) and parts (**i**), (**iv**) and (**v**) to get

$$E_{ii}(\varepsilon) = b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon)$$
$$+ \sum_{k \in \{0,N\}} (b_k[0] + b_k[1]\varepsilon + \dot{o}_k(\varepsilon))(A^*_{k,i}[-1]\varepsilon^{-1} + A^*_{k,i}[0] + o^*_{k,i}(1))$$
$$+ \sum_{k \in {}_{0,i,N}\mathbb{X}} (b_k[0] + b_k[1]\varepsilon + \dot{o}_k(\varepsilon))(A^*_{k,i}[0] + A^*_{k,i}[1]\varepsilon + o^*_{k,i}(\varepsilon)).$$

Here we can note that the terms corresponding to $k \in \{0, N\}$ are of order $O(\varepsilon^{-1})$ while all other terms are of order $O(1)$. By using the product rule and multiple summation rule for Laurent asymptotic expansions, we conclude that the asymptotic relation given in part (**vi**) also holds for $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$. □

The following theorem gives second order asymptotic expansions for stationary and conditional quasi-stationary probabilities. In particular, part (**i**) of this theorem shows that there exist limits for stationary probabilities, $\pi_i(0) = \lim_{\varepsilon \to 0} \pi_i(\varepsilon), i \in \mathbb{X}$, where $\pi_0(0) > 0, \pi_N(0) > 0$, and $\pi_i(0) = 0$ for $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$. Furthermore, part (**ii**) of the theorem shows, in particular, that there exist limits for conditional quasi-stationary probabilities, $\hat{\pi}_i(0) = \lim_{\varepsilon \to 0} \hat{\pi}_i(\varepsilon)$, $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$, where $\hat{\pi}_i(0) > 0$, $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$.

Theorem 11.4 *Assume that conditions* **A**–**G** *and* $\mathbf{H}_3$ *hold. Then:*

(**i**) *We have the following asymptotic relation for the stationary probabilities* $\pi_i(\varepsilon)$, $i \in \mathbb{X}$,

$$\pi_i(\varepsilon) = c_i[\hat{l}_i]\varepsilon^{\hat{l}_i} + c_i[\hat{l}_i + 1]\varepsilon^{\hat{l}_i+1} + o_i(\varepsilon^{\hat{l}_i+1}), \ \varepsilon \in (0, \varepsilon_0],$$

where $\hat{l}_i = I(i \neq 0, N)$, $o_i(\varepsilon^{\hat{l}_i+1})/\varepsilon^{\hat{l}_i+1} \to 0$ *as* $\varepsilon \to 0$, *and*

$$c_i[\hat{l}_i] = \frac{b_i[0]}{B_{ii}[-\hat{l}_i]} > 0, \quad c_i[\hat{l}_i + 1] = \frac{b_i[1]B_{ii}[-\hat{l}_i] - b_i[0]B_{ii}[-\hat{l}_i + 1]}{B_{ii}[-\hat{l}_i]^2},$$

where $B_{ii}[-1]$, $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$, $B_{ii}[0]$, $i \in \mathbb{X}$, *and* $B_{ii}[1]$, $i = 0, N$, *can be computed from the formulas given in Lemma 11.5.*

(ii) *We have the following asymptotic relation for the conditional quasi-stationary probabilities,* $\hat{\pi}_i(\varepsilon)$, $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$,

$$\hat{\pi}_i(\varepsilon) = \hat{c}_i[0] + \hat{c}_i[1]\varepsilon + \hat{o}_i(\varepsilon), \ \varepsilon \in (0, \varepsilon_0],$$

where $\hat{o}_i(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$ *and*

$$\hat{c}_i[0] = \frac{c_i[1]}{d[1]} > 0, \quad \hat{c}_i[1] = \frac{c_i[2]d[1] - c_i[1]d[2]}{d[1]^2},$$

where $d[l] = \sum_{j \in {}_{0,N}\mathbb{X}} c_i[l]$, $l = 1, 2$.

Proof It follows from parts **(i)** and **(vi)** in Lemma 11.5 that, for $i \in \mathbb{X}$,

$$\pi_i(\varepsilon) = \frac{b_i[0] + b_i[1]\varepsilon + \dot{o}_i(\varepsilon)}{B_{ii}[\gamma_i + \beta_i - 1]\varepsilon^{\gamma_i+\beta_i-1} + B_{ii}[\gamma_i + \beta_i]\varepsilon^{\gamma_i+\beta_i} + \dot{o}_{ii}(\varepsilon^{\gamma_i+\beta_i})}. \tag{11.98}$$

We also have $\gamma_i + \beta_i = I(i = 0) + I(i = N) = 1 - I(i \neq 0, N) = 1 - \hat{l}_i$. Using this relation for indicator functions and the division rule for Laurent asymptotic expansions in relation (11.98) we obtain the asymptotic relation given in part **(i)**.

For the proof of part **(ii)**, we first use part **(i)** for $i \in {}_{\langle 1,N-1\rangle}\mathbb{X}$ to get

$$\hat{\pi}_i(\varepsilon) = \frac{\pi_i(\varepsilon)}{\sum_{j \in {}_{0,N}\mathbb{X}} \pi_j(\varepsilon)} = \frac{c_i[1]\varepsilon + c_i[2]\varepsilon^2 + o_i(\varepsilon^2)}{\sum_{j \in {}_{0,N}\mathbb{X}}(c_j[1]\varepsilon + c_j[2]\varepsilon^2 + o_j(\varepsilon^2))},$$

and then we apply the multiple summation rule **(i)** given in Lemma 11.2 and the division rule **(iv)** given in Lemma 11.1. □

11.6.5 Asymptotic Expansions of Higher Orders for Stationary and Conditional Quasi-stationary Distributions

In Sects. 11.6.2–11.6.4, we give asymptotic expansions of the first and second orders (with length, respectively, 1 and 2) for stationary and conditional quasi-stationary distributions of perturbed semi-Markov birth-death processes, under the assumption that conditions **A**–**G** and $\mathbf{H}_i$ $(i = 1, 2, 3)$ hold.

It is readily seen from the proofs of Lemmas 11.3–11.5 and Theorems 11.2–11.4 that the perturbation conditions **D** and **E** can be weaken in the case of first order asymptotics. The asymptotic expansions of the length 2 appearing in these conditions can be replaced by the analogous asymptotic expansions of the length 1. Namely, the upper indices $1 + l_{i,\pm}$ in the sums representing these asymptotic expansions and power indices in the corresponding remainders should be, just, replaced by indices $l_{i,\pm}$.

Moreover, the method of construction of asymptotic expansions for stationary distributions and conditional quasi-stationary distribution of perturbed semi-Markov birth-death processes based on the use of operational rules for Laurent asymptotic expansions presented in Lemmas 11.1 and 11.2 let one also construct the corresponding asymptotic expansions of higher orders, with the length larger than 2. In this case, the asymptotic expansions of the length 2 appearing in the perturbation conditions **D** and **E** should be replaced by the analogous asymptotic expansions of the corresponding length larger than 2. Namely, the upper indices $1 + l_{i,\pm}$ in the sums representing these asymptotic expansions and power indices for the corresponding remainders should be formally replaced by indices $L + l_{i,\pm}$, with parameter $L > 1$. In this case, the length of the corresponding asymptotic expansions will be $L + 1$.

The algorithm for construction of the corresponding asymptotic expansions, with length $L + 1 > 2$ is absolutely analogous to those used in Theorems 11.2–11.4. The difference is that at all steps the asymptotic expansions, with length $L + 1$, are constructed for the corresponding intermediate quantities, $\Gamma_{i,j,\pm}(\varepsilon)$, etc., using operational rules for Laurent asymptotic expansions given in Lemmas 11.1 and 11.2.

This program is realised in book by Silvestrov, D. and Silvestrov, S. [41].

11.7 Numerical Examples

In this section, the results of the present paper are illustrated by numerical examples for some of the perturbed models of birth-death-type discussed in Sect. 11.2. Let us first note that each model presented in Sect. 11.2 is defined in terms of intensities for a continuous time Markov chain and the perturbation scenarios considered give intensities which are linear functions of the perturbation parameter, that is,

$$\lambda_{i,\pm}(\varepsilon) = g_{i,\pm}[0] + g_{i,\pm}[1]\varepsilon, \ i \in \mathbb{X}, \tag{11.99}$$

where the coefficients $g_{i,\pm}[l]$ depend on the model under consideration. Consequently, the higher order ($l \geq 2$) terms in (11.43) all vanish.

In order to use the algorithm based on successive reduction of the phase space, we first need to calculate the coefficients in perturbation conditions **D** and **E**. This can be done from relations (11.5), (11.6) and (11.44) by applying the operational rules for Laurent asymptotic expansions given in Lemmas 11.1 and 11.2. By relation (11.5), we have $\lambda_i(\varepsilon) = \lambda_{i,-}(\varepsilon) + \lambda_{i,+}(\varepsilon)$, so it follows immediately from (11.99) that

$$\lambda_i(\varepsilon) = g_i[0] + g_i[1]\varepsilon,\ i \in \mathbb{X}, \tag{11.100}$$

where $g_i[l] = g_{i,-}[l] + g_{i,+}[l]$, $l = 0, 1$. From (11.6), (11.99), (11.100) and Lemma 11.1 we deduce the following asymptotic series expansions, for $i \in \mathbb{X}$,

$$\begin{aligned} p_{i,\pm}(\varepsilon) &= \frac{\lambda_{i,\pm}(\varepsilon)}{\lambda_i(\varepsilon)} = \frac{g_{i,\pm}[0] + g_{i,\pm}[1]\varepsilon}{g_i[0] + g_i[1]\varepsilon} \\ &= \sum_{l=l_{i,\pm}}^{1+l_{i,\pm}} a_{i,\pm}[l]\varepsilon^l + o_{i,\pm}(\varepsilon^{1+l_{i,\pm}}). \end{aligned} \tag{11.101}$$

The expansion (11.101) exists and its coefficients can be calculated from the division rule for asymptotic expansions. Then, using (11.44), (11.100), (11.101) and Lemma 11.1, the following asymptotic series expansions can be constructed, for $i \in \mathbb{X}$,

$$e_{i,\pm}(\varepsilon) = \frac{p_{i,\pm}(\varepsilon)}{\lambda_i(\varepsilon)} = \sum_{l=l_{i,\pm}}^{1+l_{i,\pm}} b_{i,\pm}[l]\varepsilon^l + \dot{o}_{i,\pm}(\varepsilon^{1+l_{i,\pm}}). \tag{11.102}$$

Once the coefficients in the expansions (11.101) and (11.102) have been calculated, we can use the algorithm described in Sect. 11.6, in order to construct asymptotic expansions for stationary and conditional quasi-stationary probabilities.

The remainder of this section is organised as follows. In Sect. 11.7.1 we illustrate our results with numerical calculations for the perturbed models of population genetics discussed in Sect. 11.2.3. We first consider an example where condition $\mathbf{H}_1$ holds and then an example where condition $\mathbf{H}_3$ is satisfied. Numerical examples for the perturbed model of epidemics presented in Sect. 11.2.2 are discussed in Sect. 11.7.2. This provides an example where condition $\mathbf{H}_2$ holds. All illustrations for the numerical examples are placed in a special subsection at the end of this section for convenience.

11.7.1 Numerical Examples for Perturbed Models of Population Genetics

Recall that the perturbation conditions for the model in Sect. 11.2.3 are formulated in terms of the mutation parameters as

$$U_1(\varepsilon) = C_1 + D_1\varepsilon, \quad U_2(\varepsilon) = C_2 + D_2\varepsilon. \tag{11.103}$$

Additionally, the model depends on the size $N/2$ of the population and the selection parameters S_1 and S_2 which are assumed to be independent of ε. Thus, there are in total seven parameters to choose.

In our first example, we choose the following values for the parameters: $N = 100$, $C_1 = C_2 = 5$, $D_1 = 0$, $D_2 = N$ and $S_1 = S_2 = 0$. Recall that the mutation probabilities are related to the mutation parameters by $u_1(\varepsilon) = U_1(\varepsilon)/N$ and $u_2(\varepsilon) = U_2(\varepsilon)/N$. It follows from (11.103) that $u_1(\varepsilon) = 0.05$ and $u_2(\varepsilon) = 0.05 + \varepsilon$. Thus, in the limiting model, a chosen allele mutates with probability 0.05 for both types A_1 and A_2. In this case, we have no absorbing states which means that condition $\mathbf{H}_1$ holds.

Since we have no selection, the stationary distribution for the limiting model will be symmetric around state 50. The perturbation parameter ε can be interpreted as an increase in the probability that a chosen allele of type A_2 mutates to an allele of type A_1. Increasing the perturbation parameter will shift the mass of the stationary distribution to the right.

With model parameters given above, we first used relations (11.20), (11.21), (11.22), (11.24), and (11.26) to calculate the coefficients in (11.99) for the intensities. Then, these coefficients were used to compute the coefficients in the perturbation conditions **D** and **E** as described above. After this, we used the algorithm outlined in Sect. 11.6 to calculate the asymptotic expansions for the stationary distribution given in Theorem 11.2, with parameters $L = 0, 1$, i.e., lengths $L + 1 = 1, 2$. Moreover, we also computed the analogous asymptotic expansions, with parameter $L = 2, 3$, i.e., lengths $L + 1 = 3, 4$, using the higher order variant of the corresponding algorithm described in Sect. 11.6.5. Approximations for the stationary distribution based on these expansions were obtained by approximating the corresponding remainders by zero.

Let us first compare our approximations with the exact stationary distribution for some particular values of the perturbation parameter. Figure 11.1a shows the stationary distribution for the limiting model ($\varepsilon = 0$) and, as already mentioned above, we see that it is symmetric around state 50. The stationary distribution for the model with $\varepsilon = 0.01$ and the approximation corresponding to $L = 1$ are shown in Fig. 11.1b. Here, the approximation seems the match the exact distribution very well. The approximation for $L = 2$ is not included here since it will not show any visible difference from the exact stationary distribution. In Fig. 11.1c, d, corresponding to the models where $\varepsilon = 0.02$ and $\varepsilon = 0.03$, respectively, we also include the approximations for $L = 2$. As expected, the approximations for the stationary distribution get worse as the perturbation parameter increases. However, it seems that even for higher values of the perturbation parameter, some parts of our approximations fit better to the exact stationary distribution. In this example, it seems that the approximations are in general better for states that belong to the right part of the distribution.

In order to illustrate that the quality of the approximations differs depending on which states we consider, let us compare the stationary probabilities for the states 40 and 80. The stationary probabilities of these two states are approximately of the same magnitude and we can compare them in plots with the same scale on both the horizontal and the vertical axes. Figure 11.2a shows the stationary probability for state 40 as a function of the perturbation parameter and its approximations for $L = 1, 2, 3$. The corresponding quantities for state 80 are shown in Fig. 11.2b where we have omitted the approximation for $L = 3$ since the approximation is very good

already for $L = 2$. When $L = 2$, the approximation for state 80 is clearly better compared to the approximation for state 40.

Another point illustrated by Figs. 11.1 and 11.2 is that for a fixed value of the perturbation parameter, the quality of an approximation based on a higher order asymptotic expansion is not necessarily better. For instance, in Fig. 11.2a we see that for $\varepsilon \in [0.04, 0.05]$ the approximations for $L = 1$ is better compared to both $L = 2$ and $L = 3$. However, asymptotically as $\varepsilon \to 0$, the higher order approximations are better. For example, we see in Fig. 11.2a that when $\varepsilon \in [0, 0.02]$ the approximations for $L = 3$ are the best.

Let us now consider a second example for the perturbed model of population genetics. We now choose the parameters as follows: $N = 100$, $C_1 = C_2 = 0$, $D_1 = D_2 = N$ and $S_1 = S_2 = 0$. In this case, both types of mutations have the same probabilities and are equal to the perturbation parameter, that is, $u_1(\varepsilon) = u_2(\varepsilon) = \varepsilon$. This means that both boundary states will be asymptotically absorbing, so condition $\mathbf{H}_3$ holds. In this case, we calculated the asymptotic expansions for the stationary and conditional quasi-stationary stationary distribution, given in Theorem 11.4.

Let us illustrate the numerical results for conditional quasi-stationary distributions. Figure 11.3a shows the conditional quasi-stationary distribution for $\varepsilon = 0.005$ and some of its approximations. Since it is quite hard to see the details near the boundary states for this plot, we also show the same curves restricted to the states 1–20 in Fig. 11.3b. As in the previous example, it can be seen that the qualities of the approximations differ between the states. In this case, we see that the approximations for states close to the boundary are not as good as for interior states. Similar type of behaviour also appears for different choices of the selection parameters S_1 and S_2. We omit the plots showing this since they do not contribute with more understanding of the model.

Let us instead study the limiting conditional quasi-stationary distributions (11.51) for some different values of the selection parameters S_1 and S_2. These types of distributions are interesting in their own right and are studied, for instance, by Allen and Tarnita [1], where they are called *rare-mutation dimorphic distributions*. In our example, if mutations are rare (i.e., ε is very small), the probabilities of such a distribution can be interpreted as the likelihoods for different allele frequencies to appear during periods of competition which are separated by long periods of fixation.

Figure 11.4a shows the limiting conditional quasi-stationary distribution in the case $S_1 = S_2 = 0$, that is, for a selectively neutral model. Now, let the selection parameters be given by $S_1 = 10$ and $S_2 = -10$. In this case, the gene pairs with genotypes A_1A_1, A_2A_2 and A_1A_2 have survival probabilities approximately equal to 0.37, 0.30 and 0.33, respectively. Thus, allele A_1 has a selective advantage and this is reflected in Fig. 11.4b where the limiting conditional quasi-stationary distribution is shown in this case. The mass of the distribution is now shifted to the right compared to a selectively neutral model. Next, we take the selection parameters as $S_1 = S_2 = 10$ which implies that gene pairs with genotypes A_1A_1, A_2A_2 and A_1A_2 have survival probabilities approximately equal to 0.345, 0.345 and 0.31, respectively. This means that we have a model with underdominance and we see in Fig. 11.4c that the limiting conditional quasi-stationary distribution then has more of its mass near the boundary

compared to a selectively neutral model. Finally, we set the selection parameters as $S_1 = S_2 = -10$. Then, gene pairs with genotypes A_1A_1, A_2A_2 and A_1A_2 have survival probabilities approximately equal to 0.32, 0.32 and 0.36, respectively. This gives us a model with overdominace or balancing selection and in this case we see in Fig. 11.4d that the limiting conditional quasi-stationary distribution has more mass concentrated to the interior states compared to a selectively neutral model.

11.7.2 Numerical Examples for Perturbed Epidemic Models

In our last numerical example, we consider the perturbed epidemic model described in Sect. 11.2.2. Recall from the variant of condition $\mathbf{H}_2$ given in this subsection that the contact rate ν for each individual and the group of infected individuals outside the population is considered as a perturbation parameter, that is, $\nu = \nu(\varepsilon) = \varepsilon$. In this case, state 0 is asymptotically absorbing which means that condition $\mathbf{H}_2$ holds.

It follows directly from (11.15) and (11.16) that the intensities of the Markov chain describing the number of infected individuals are linear functions of ε given by $\lambda_{i,+}(\varepsilon) = \lambda i(1 - i/N) + (N - i)\varepsilon$, $\lambda_{i,-}(\varepsilon) = \mu i$, $i \in \mathbb{X}$. In this model, we only have three parameters to choose: N, λ, and μ. As in the previous examples, let us take $N = 100$ which here corresponds to the size of the population. Furthermore, we let $\mu = 1$ so that the expected time for an infected individual to be infectious is equal to one time unit. Numerical illustrations will be given for the cases where $\lambda = 0.5$ and $\lambda = 1.5$. For the limiting model, we have in the former case that the basic reproduction ratio $R_0 = 0.5$ and in the latter case $R_0 = 1.5$. The properties of the model are quite different depending on which of these two cases we consider. For the two choices of model parameters given above, we calculated asymptotic expansions for stationary and conditional quasi-stationary distributions given in Theorem 11.3.

Let us first compare the limiting conditional quasi-stationary distributions in (11.49). Figure 11.5a shows this distribution for the case where $\lambda = 0.5$ and $\mu = 1$ and in Fig. 11.5b it is shown for the case where $\lambda = 1.5$ and $\mu = 1$. In the former case, the limiting conditional quasi-stationary distribution has most of its mass concentrated near zero and in the latter case the distribution has a shape which resembles a normal curve and most of its mass is distributed on the states between 0 and 60.

We can also study plots of the type given in Figs. 11.1, 11.2 and 11.3. Also in this example, intervals for the perturbation parameter, where the approximations are good, depend on which state is considered. In this case, states close to zero are more sensitive to perturbations. Let us here just show two of the plots for illustration. For the model with $\lambda = 1.5$ and $\mu = 1$, Fig. 11.6a shows the conditional quasi-stationary distribution for $\varepsilon = 0.02$ and the corresponding approximations for $L = 1$ and $L = 2$. For the same model parameters, the quasi-stationary probability for state 10 is shown in Fig. 11.6b as a function of the perturbation parameter together with some of its approximations.

Finally, let us compare the stationary probabilities for state 0. Note that, despite that the limiting conditional quasi-stationary distribution is very different depending

on whether $R_0 = 0.5$ or $R_0 = 1.5$ for the model with $\varepsilon = 0$, the limiting stationary distribution is concentrated at state 0 in both these cases. Figure 11.7a shows the stationary probability of state 0 as a function of the perturbation parameter and some of its approximations in the case where $\lambda = 0.5$ and $\mu = 1$. The corresponding quantities for the case where $\lambda = 1.5$ and $\mu = 1$ are shown in Fig. 11.7b.

Qualitatively the plots show approximately the same behavior, but note that the scales on the horizontal axes are very different. We see that the stationary probability of state 0 for the limiting model is much more sensitive to perturbations in the case where $R_0 = 1.5$. It follows from (11.52) that this is due to fact that the expected time $E_{10}(\varepsilon)$ for the infection to (temporarily) die out after one individual gets infected, is much larger for the model with $R_0 = 1.5$.

Illustrations for Numerical Examples

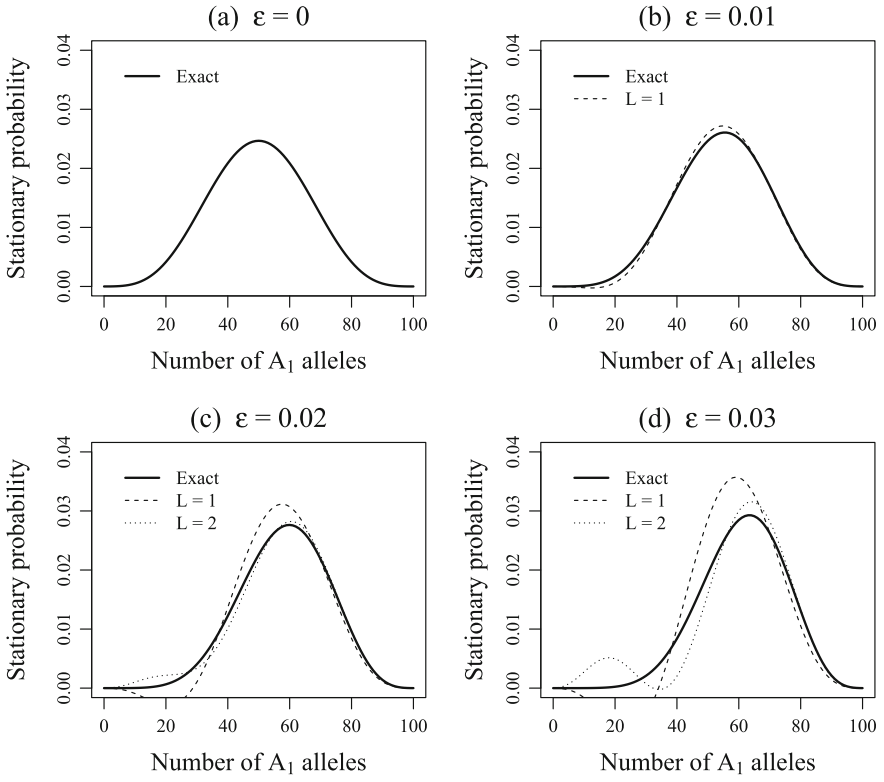

Fig. 11.1 Comparison of the stationary distribution $\pi_i(\varepsilon)$ and some of its approximations for the population genetic example of Sect. 11.2.3. The plots are functions of the number of A_1 alleles i, for different values of the perturbation parameter ε, with $N = 100$, $C_1 = C_2 = 5$, $D_1 = 0$, $D_2 = N$ and $S_1 = S_2 = 0$

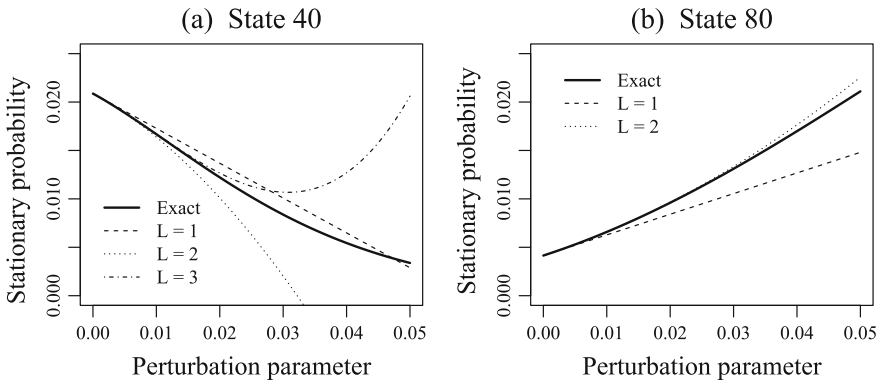

Fig. 11.2 Comparison of stationary probabilities $\pi_i(\varepsilon)$ for states $i = 40$ and $i = 80$ and some of its approximations considered as a function of the perturbation parameter ε. The model is based on the population genetic example of Sect. 11.2.3, with the same parameter values as in Fig. 11.1

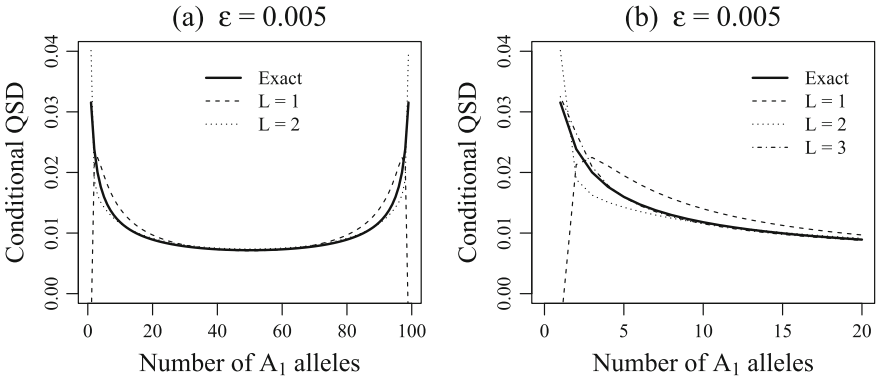

Fig. 11.3 The conditional quasi-stationary distribution $\hat{\pi}_i(\varepsilon)$ and some of its approximations for the population genetic example of Sect. 11.2.3. The plots are functions of the number of A_1 alleles i, with the perturbation parameter $\varepsilon = 0.005$ fixed. Plot **a** shows the distribution for all states while plot **b** is restricted to states 1–20. The parameter values of the model are $N = 100$, $C_1, C_2 = 0$, $D_1, D_2 = N$ and $S_1, S_2 = 0$

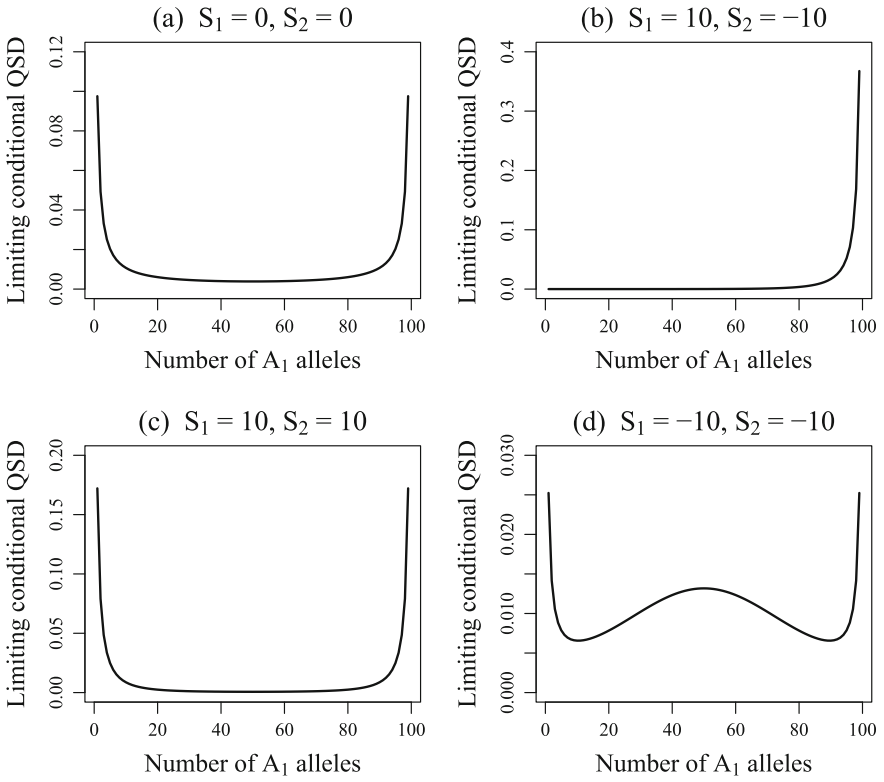

Fig. 11.4 Plots of the limiting conditional quasi-stationary distribution $\hat{\pi}_i(0)$ for the population genetic example of Sect. 11.2.3, as a function of the number of A_1-alleles i, for different values of the selection parameters. The model parameters N, C_1, C_2, D_1 and D_2 are the same as in Fig. 11.3. Note that the scales of the vertical axes differ between the plots

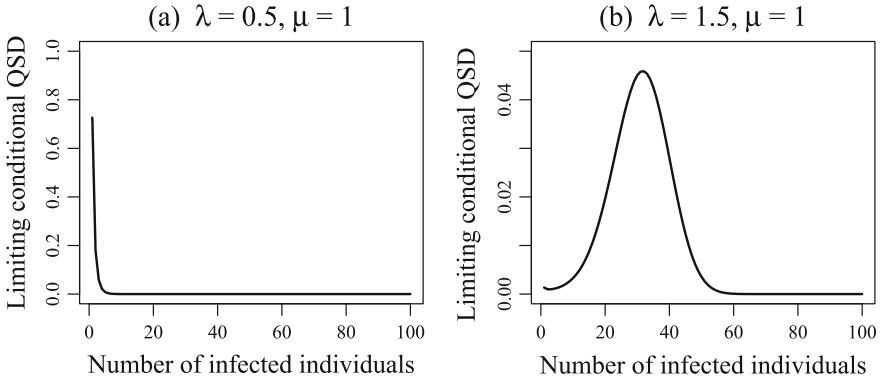

Fig. 11.5 Comparison of the limiting conditional quasi-stationary distribution $\tilde{\pi}_i(0)$ for the epidemic model of Sect. 11.2.2, as a function of the number of infected individuals i, for a population of size $N = 100$ with recovery rate $\mu = 1$. The force of infection parameter is $\lambda = 0.5$ in **a** and $\lambda = 1.5$ in **b**. Note that the scales of the vertical axes differ between the two plots

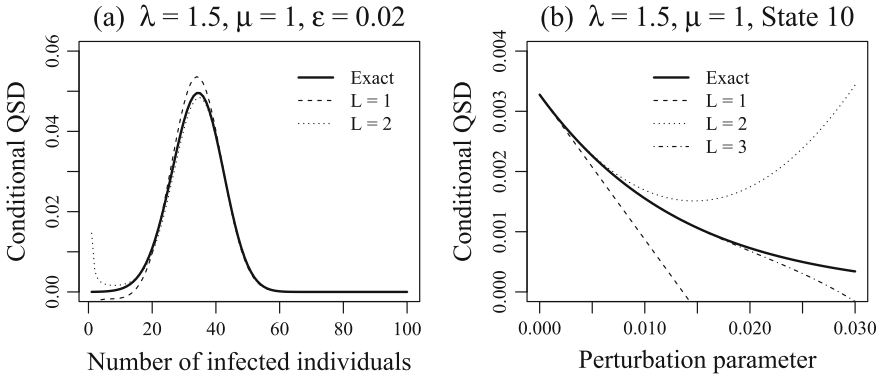

Fig. 11.6 Conditional quasi-stationary probabilities $\tilde{\pi}_i(\varepsilon)$ and some approximations for the epidemic model of Sect. 11.2.2, with $N = 100$, $\lambda = 1.5$ and $\mu = 1$. Note that the horizontal axes in the two plots represent different quantities; the number of infected individuals i in **a** and the perturbation parameter ε in **b**

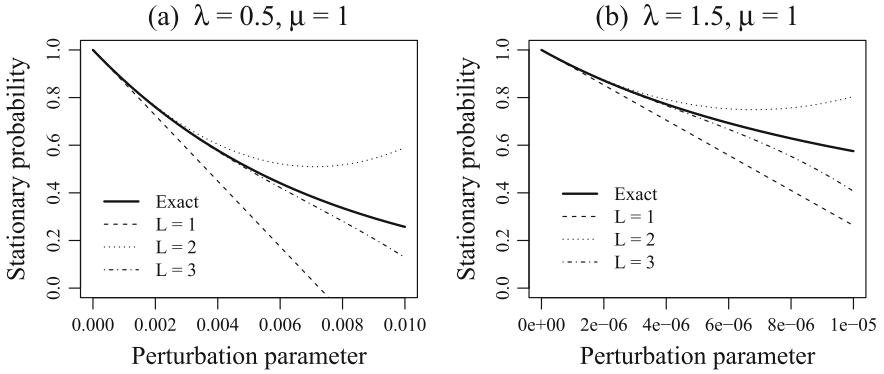

Fig. 11.7 Comparison of the stationary probability $\pi_i(\varepsilon)$ of state $i = 0$ as a function of the perturbation parameter ε for the epidemic model of Sect. 11.2.2 when $N = 100$, $\mu = 1$, and the contact rate parameter equals **a** $\lambda = 0.5$ and **b** $\lambda = 1.5$. Note that the scales of the horizontal axes differ between the two plots

11.8 Discussion

The present paper is devoted to studies of asymptotic expansions for stationary and conditional quasi-stationary distributions for perturbed semi-Markov birth-death processes. We employ the algorithms of sequential phase space reduction for perturbed semi-Markov processes combined with techniques of Laurent asymptotic expansions developed in the recent works by Silvestrov, D. and Silvestrov, S. [39–41], and apply them to semi-Markov birth-death processes. In this model, the proposed algorithms of phase space reduction preserve the birth-death structure for reduced semi-Markov processes. This made it possible to get, in the present paper, explicit

formulas for coefficients in the corresponding asymptotic expansions of the first and the second orders, for stationary and conditional quasi-stationary distributions of perturbed semi-Markov birth-death processes.

The above results are applied to three types of perturbed models from biology; population dynamics, epidemic models and models of population genetics. We supplement theoretical results by computations, illustrating numerical accuracy of the corresponding asymptotic expansions for stationary and quasi-stationary distributions of varying form. Even though exact expressions for the (quasi-)stationary distributions of these biological models are available, the asymptotic expansions may still be preferable when the state space is large and (quasi-)stationary distributions are computed for several values of the perturbation parameter, since only the first coefficients of the appropriate Laurent expansions are needed.

It should be mentioned that the semi-Markov setting is an adequate and necessary element of the proposed method. Even in the case where the initial birth-death-type process is a discrete or continuous time Markov chain, the time-space screening procedure of phase space reduction results in a semi-Markov birth-death process, since times between sequential hitting of the reduced space by the initial process have distributions which can differ from geometrical or exponential ones.

Also, the use of Laurent asymptotic expansions for expectations of sojourn times of perturbed semi-Markov processes is a necessary element of the proposed method. Indeed, even when expectations of sojourn times for all states of the initial semi-Markov birth-death process are asymptotically bounded and represented by Taylor asymptotic expansions, the exclusion of an asymptotically absorbing state from the initial phase space can generate states with asymptotically unbounded expectations of sojourn times represented by Laurent asymptotic expansions, for the reduced semi-Markov birth-death processes.

Several extensions of our work are possible. We have considered semi-Markov processes defined on a finite and linearly ordered state space $\mathbb{X}$, that is a subset of a one-dimensional lattice. We also confined ourselves to processes of birth-death type, where only jumps to neighboring states are possible.

For population dynamics models, it is noted by Lande, Engen and Saether [26] that one needs to go beyond birth-death processes though and incorporate larger jumps in order to account for a changing environment. State spaces that are subsets of higher-dimensional lattices are of interest in a number of applications, for instance SIR-models of epidemic spread where some recovered individuals get immune, Nåsell [32], population genetic models with two sexes, Moran [28], Hössjer and Tyvand [16], and population dynamics or population genetics models with several species or subpopulations, see Lande, Engen and Saether [26], Hössjer et al. [15] and references therein. It is an interesting topic of further research to apply the methodology of this paper to such models.

The method of sequential phase space reduction proposed in this paper can be applied to get asymptotic expansions for high order power and mixed power-exponential moments of hitting times and, in sequel, for more complex

quasi-stationary distributions (given by relation (11.41)) for nonlinearly perturbed semi-Markov birth-death processes and, thus, for models of population dynamics, epidemic spread and population genetics, which are the objects of interest in the present paper. We hope to present such results in the future.

References

1. Allen, B., Tarnita, C.E.: Measures of success in a class of evolutionary models with fixed population size and structure. J. Math. Biol. **68**, 109–143 (2014)
2. Allen, L.J.S., Burgin, A.M.: Comparison of deterministic and stochastic SIS and SIR models in discrete time. Math. Biosci. **163**, 1–33 (2000)
3. Avrachenkov, K.E., Filar, J.A., Howlett, P.G.: Analytic Perturbation Theory and Its Applications, xii+372 pp. SIAM, Philadelphia (2013)
4. Bini, D.A., Latouche, G., Meini, B.: Numerical Methods for Structured Markov Chains. Numerical Mathematics and Scientific Computation. Oxford Science Publications, xii+327 pp. Oxford University Press, New York (2005)
5. Cavender, J.A.: Quasi-stationary distributions of birth-and-death processes. Adv. Appl. Probab. **10**, 570–586 (1978)
6. Collet, P., Martínez, S., San Martín, J.: Quasi-Stationary Distributions. Markov Chains, Diffusions and Dynamical Systems. Probability and its Applications, xvi+280 pp. Springer, Heidelberg (2013)
7. Crow, J.F., Kimura, M.: An Introduction to Population Genetics Theory, xiv+591 pp. Harper and Row Publishers, New York (1970) (The Blackburn Press, Caldwell, NJ, pp. 608 (2009))
8. Durrett, R.: Probability Models for DNA Sequence Evolution, xii+431 pp. Springer, New York (2008) (2nd revised edition of Probability Models for DNA Sequence Evolution. Springer, New York, viii+240 pp. (2002))
9. Ewens, W.J.: Mathematical Population Genetics I. Theoretical Introduction, xx+417 pp. Springer, New York (2004) (2nd revised edition of Mathematical Population Genetics. Biomathematics, vol. 9, xii+325 pp. Springer, Berlin (1979))
10. Feller, W.: Die Grundlagen Volterraschen Theorie des Kampes ums Dasein in wahrscheinlichkeitsteoretischer Behandlung. Acta Biotheor. **5**, 11–40 (1939)
11. Feller, W.: An Introduction to Probability Theory and Its Applications, xviii+509 pp. Wiley, New York (1968) (3rd edition of An Introduction to Probability Theory and Its Applications, vol. I, xii+419 pp. Wiley, New York (1950))
12. Gilpin, M.E., Ayala, F.J.: Global models of growth and competition. Proc. Natl. Acad. Sci. USA **70**, 3590–3593 (1973)
13. Gyllenberg, M., Silvestrov, D.S.: Quasi-Stationary Phenomena in Nonlinearly Perturbed Stochastic Systems. De Gruyter Expositions in Mathematics, vol. 44, ix+579 pp. Walter de Gruyter, Berlin (2008)
14. Hethcote, H.W.: The mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
15. Hössjer, O., Olsson, F., Laikre, L., Ryman, N.: A new general analytical approach for modeling patterns of genetic differentiation and effective size of subdivided populations over time. Math. Biosci. **258**, 113–133 (2014)
16. Hössjer, O., Tyvand, P.: A monoecious and diploid Moran model of random mating. J. Theor. Biol. **394**, 182–196 (2016)
17. Hössjer, O., Tyvand, P.A., Miloh, T.: Exact Markov chain and approximate diffusion solution for haploid genetic drift with one-way mutation. Math. Biosci. **272**, 100–112 (2016)
18. Jacquez, J.A., O'Neill, P.: Reproduction numbers and thresholds in stochastic epidemic models I. Homogenous populations. Math. Biosci. **107**, 161–186 (1991)
19. Jacquez, J.A., Simon, C.P.: The stochastic SI model with recruitment and deaths I. Comparisons with the closed SIS model. Math. Biosci. **117**, 77–125 (1993)

20. Karlin, S., McGregor, J.: On a genetics model of Moran. Proc. Camb. Philos. Soc. **58**, 299–311 (1962)
21. Kendall, D.G.: Stochastic processes and population growth. J. R. Stat. Soc. Ser. B **11**, 230–264 (1949)
22. Konstantinov, M., Gu, D.W., Mehrmann, V., Petkov, P.: Perturbation Theory for Matrix Equations. Studies in Computational Mathematics, vol. 9, xii+429 pp. North-Holland, Amsterdam (2003)
23. Korolyuk, V.S., Korolyuk, V.V.: Stochastic Models of Systems. Mathematics and its Applications, vol. 469, xii+185 pp. Kluwer, Dordrecht (1999)
24. Koroliuk, V.S., Limnios, N.: Stochastic Systems in Merging Phase Space, xv+331 pp. World Scientific, Singapore (2005)
25. Kryscio, R.J., Lefévre, C.: On the extinction of the S-I-S stochastic logistic epidemic. J. Appl. Probab. **27**, 685–694 (1989)
26. Lande, R., Engen, S., Saether, B.-E.: Stochastic Population Dynamics in Ecology and Conservation. Oxford Series and Ecology and Evolution, x+212 pp. Oxford University Press, Oxford (2003)
27. Moran, P.A.P.: Random processes in genetics. Proc. Camb. Philos. Soc. **54**, 60–71 (1958)
28. Moran, P.A.P.: A general theory of the distribution of gene frequencies I. Overlapping generations. Proc. Camb. Philos. Soc. **B149**, 102–112 (1958)
29. Nåsell, I.: The quasi-stationary distribution of the closed endemic SIS model. Adv. Appl. Probab. **28**, 895–932 (1996)
30. Nåsell, I.: On the quasi-stationary distribution of the stochastic logistic epidemic. Math. Biosci. **156**, 21–40 (1999)
31. Nåsell, I.: Extinction and quasi-stationarity of the Verhulst logistic model. J. Theor. Biol. **211**, 11–27 (2001)
32. Nåsell, I.: Stochastic models of some endemic infections. Math. Biosci. **179**, 1–19 (2002)
33. Nåsell, I.: Moment closure and the stochastic logistic model. Theor. Popul. Biol. **63**, 159–168 (2003)
34. Nåsell, I.: Extinction and Quasi-Stationarity in the Stochastic Logistic SIS Model. Lecture Notes in Mathematics, Mathematical Biosciences Subseries, xii+199 pp. Springer, Heidelberg (2011)
35. Pearl, R.: The growth of populations. Q. Rev. Biol. **2**, 532–548 (1927)
36. Petersson, M.: Asymptotics for quasi-stationary distributions of perturbed discrete time semi-Markov processes. In: Silvestrov, S., Rančić, M. (eds.) Engineering Mathematics II. Algebraic, Stochastic and Analysis Structures for Networks, Data Classification and Optimization. Springer Proceedings in Mathematics and Statistics, vol. 179, pp. 131–149. Springer, Heidelberg (2016)
37. Petersson, M.: Perturbed discrete time stochastic models. Doctoral dissertation, Stockholm University (2016)
38. Silvestrov, D., Manca, R.: Reward algorithms for semi-Markov processes. Methodol. Comput. Appl. Probab. **19**(4), 1191–1209 (2017)
39. Silvestrov, D., Silvestrov, S.: Asymptotic expansions for stationary distributions of perturbed semi-Markov processes. In: Silvestrov, S., Rančić, M. (eds.) Engineering Mathematics II. Algebraic, Stochastic and Analysis Structures for Networks, Data Classification and Optimization. Springer Proceedings in Mathematics and Statistics, vol. 179, pp. 151–222. Springer, Cham (2016)
40. Silvestrov, D., Silvestrov, S.: Asymptotic expansions for stationary distributions of nonlinearly perturbed semi-Markov processes 1, 2. Methodol. Comput. Appl. Probab. 20 (2017). Part 1: https://doi.org/10.1007/s11009-017-9605-0, Part 2: https://doi.org/10.1007/s11009-017-9607-y
41. Silvestrov, D., Silvestrov, S.: Nonlinearly Perturbed Semi-Markov Processes. Springer Briefs in Probability and Mathematical Statistics, xiv+143 pp. Springer, Berlin (2017)
42. Stewart, G.W.: Matrix Algorithms. Vol. I. Basic Decompositions, xx+458 pp. SIAM, Philadelphia (1998)

43. Stewart, G.W.: Matrix Algorithms. Vol. II. Eigensystems, xx+469 pp. SIAM, Philadelphia (2001)
44. Tsoularis, A., Wallace, J.: Analysis of logistic growth models. Math. Biosci. **179**, 21–55 (2002)
45. Verhulst, P.F.: Notice sur la loi que la population suit dans son accroissement. Corr. Math. Phys. **10**, 113–121 (1838)
46. Weiss, G.H., Dishon, M.: On the asymptotic behavior of the stochastic and deterministic models of an epidemic. Math. Biosci. **11**, 261–265 (1971)
47. Whittle, P.: On the use of the normal approximation in the treatment of stochastic processes. J. R. Stat. Soc. Ser. B **19**, 268–281 (1957)
48. Yin, G.G., Zhang, Q.: Discrete-Time Markov Chains. Two-Time-Scale Methods and Applications. Stochastic Modelling and Applied Probability, vol. 55, xix+348 pp. Springer, New York (2005)
49. Yin, G.G., Zhang, Q.: Continuous-Time Markov Chains and Applications. A Two-Time-Scale Approach. Stochastic Modelling and Applied Probability, vol. 37, 2nd edn, xxii+427 pp. Springer, New York (2013) (An extended variant of the first 1998 edition)

Chapter 12
Phase-Type Distribution Approximations of the Waiting Time Until Coordinated Mutations Get Fixed in a Population

Ola Hössjer, Günter Bechly and Ann Gauger

Abstract In this paper we study the waiting time until a number of coordinated mutations occur in a population that reproduces according to a continuous time Markov process of Moran type. It is assumed that any individual can have one of $m + 1$ different types, numbered as $0, 1, \ldots, m$, where initially all individuals have the same type 0. The waiting time is the time until all individuals in the population have acquired type m, under different scenarios for the rates at which forward mutations $i \to i + 1$ and backward mutations $i \to i - 1$ occur, and the selective fitness of the mutations. Although this waiting time is the time until the Markov process reaches its absorbing state, the state space of this process is huge for all but very small population sizes. The problem can be simplified though if all mutation rates are smaller than the inverse population size. The population then switches abruptly between different fixed states, where one type at a time dominates. Based on this, we show that phase-type distributions can be used to find closed form approximations for the waiting time law. Our results generalize work by Schweinsberg [60] and Durrett et al. [20], and they have numerous applications. This includes onset and growth of cancer for a cell population within a tissue, with type representing the severity of the cancer. Another application is temporal changes of gene expression among the individuals in a species, with type representing different binding sites that appear in regulatory sequences of DNA.

Keywords Coordinated mutations · Fixed state population · Moran model Phase-type distribution · Waiting time

O. Hössjer (✉)
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
e-mail: ola@math.su.se

G. Bechly · A. Gauger
Biologic Institute, 16310 NE 80th Street, Redmond, WA 98052, USA
e-mail: bechly@mac.com

A. Gauger
e-mail: ann.gauger@gmail.com

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_12

12.1 Introduction

A central problem of population genetics is to calculate the probability that a new germline point mutation survives and spreads from one individual to the rest of the population. This fixation probability depends not only on the selective fitness of the mutant compared to the wildtype variant, but also on the size of the population. Fisher [22, 23], Haldane [30], and Wright [67, 69] derived formulas for the fixation probability of a homogeneous one-sex or two-sex population without any subdivision. Their results were generalized by Kimura [34, 35], who formulated the fixation probability as a solution of Kolmogorov's backward equation. More recently, Lambert [41] gave a unified continuous branching process framework for calculating fixation probabilities of different population models. However, in order to know how fast genetic changes occur in a population, it is not only important to know the fixation probability, but also how long it takes for a surviving mutation to spread. This can be quantified in terms of the expected time until fixation, and for a homogeneous population this expected time was derived by Kimura and Ohta [38], Maruyama and Kimura [47, 48], and Kimura [36].

The above mentioned results have been generalized in different directions. First, a number of authors have analyzed fixation probabilities or the time to fixation of one single point mutation for models with geographic subdivision (Maruyama [46], Slatkin [61], Barton [3], Whitlock [65], Greven et al. [28]). Second, others have studied the waiting time until a more general type of DNA target gets fixed in a population, a process which involves several point mutations. This target could, for instance, be a double mutant at two loci with or without recombination (Bodmer [10], Christiansen et al. [14]). Another target is a subset of all possible DNA sequences at a number of tightly linked nucleotides. The evolutionary process then becomes a random walk on a fitness landscape of DNA strings, until the target set is reached (Gillespie [26], Chatterjee et al. [13]). This DNA string could, for instance, represent a regulatory region of a gene, and the target may consist of all sequences that contain a certain binding site of length 6–10 nucleotides, to which a transcription factor attaches and affects the expression of the gene (Stone and Wray [63], MacArthur and Brockfield [45], Yona et al. [70]). The waiting time until the new binding site arrives and gets fixed not only depends on the mutation rate, its selective advantage, and the size of the population, but also on the length of the regulatory region and the binding site, see for instance Durrett and Schmidt [18], Behrens and Vingron [7], Behrens et al. [8], Nicodéme [51], Tuğrul et el. [64], and Sanford et al. [56, 57].

Third, if several point mutations are required to reach a target that represents a complex adaption, these mutations must be coordinated in some way. For instance, it has been known for long that it is very difficult for several coordinated mutations to spread and get fixed if the intermediate states convey a selective disadvantage. In order for this to happen, the population has to be small or the mutations have to arrive fairly close in time. Wright's shifting balance theory (Wright [67, 68]) is an early attempt to explain this through geographic subdivision, where the coordinated mutations first occur and get fixed locally, before they spread to other subpopulations.

Kimura [37] considered a diploid model, and used a diffusion approach in order to find the expected waiting time until two coordinated mutations get fixed in the population, when each mutation by itself is negatively selected for, and the two loci are tightly linked or have a small recombination fraction between them. He approximated the two-dimensional process for the frequencies of the two mutant genes by a simpler, one-dimensional process. Stephan [62] generalized Kimura's model by allowing the two pathways towards the double mutation to have different mutation rates and selective disadvantages. Phillips [53] studied waiting times for two coordinated mutations to appear, using a somewhat similar model. He applied the solution to the first local phase of Wright's shifting-balance theory, and argued that this phase dominates the total waiting time until global fixation occurs.

The waiting time problem for coordinated mutations has several applications. It is widely believed, for instance, that many types of cancer occur when several somatic mutations spread in a population of cells within a tissue (Knudson [39]). This has been analyzed mathematically by Komarova et al. [40], Iwasa et al. [32, 33], Nowak [52], and Schinazi [58, 59]. A second related application is immune system response, where coordinated somatic mutations are triggered in reaction to certain antigens (Radmacher et al. [54]). A third application is to analyze the waiting time until multiple germline mutations arrive in duplicate genes in order to make them functional (Behe and Snoke [5, 6], Lynch [43]). A fourth application is coordinated germline mutations in regulatory regions, where changes at two different binding sites have to occur in a given order (Carter and Wagner [11], Durrett and Schmidt [19]). A fifth application is coordinated mutations in bacterial populations, where each surviving mutant gives rise to a daughter population that grows at an exponential rate (Axe [2]).

It is challenging to define a population genetic model that gives explicit formulas for the waiting time until several coordinated mutations occur. The reason is that such a model has to incorporate random gene frequency variation in terms of genetic drift, apart from selection and mutation. It is therefore necessary to study the time dynamics of the population's genetic composition by means of a stochastic process, and with at least two coordinated mutations, the state space of this process gets huge for all but very small populations sizes. Under certain assumptions the problem can be simplified though. For models with two coordinated mutations, this has been done in the above mentioned papers by Komarova et al. [40], Iwasa et al. [32, 33], and Durrett and Schmidt [19]. More recently, Schweinsberg [60] and Durrett et al. [20] obtained the asymptotic distribution for the waiting time until an arbitrary number m of coordinated mutations occur, when the intermediate alleles are neutral and no backward mutations are allowed. Their models have been used and extended by Lynch and Abegg [44], in order to study the waiting time until complex adaptive mutations are fixed. This work has been criticized by Axe [2], who argued that backward mutations should be included in models of complex adaptations.

The purpose of this paper is to generalize the framework of Schweinsberg [60] and Durrett et al. [20]. We derive asymptotic properties of the waiting time distribution until an arbitrary number m of coordinated mutations appear and the last one of them gets fixed, in a large population without any type of subdivision.

The mutations are allowed to have different selective fitness and mutation rates, and backward mutations are possible. The mutation probabilities are assumed to be smaller than the inverse population size, so that the genetic composition of the population changes rapidly between fixation of different genetic variants. This fixed state population model (Komarova et al. [40], Tuğrul et al. [64]) is conveniently modeled by a continuous time Markov process with a finite state space; the wildtype genetic variant and the m mutants. It is shown that asymptotically, the time until the mth mutant gets fixed in the population, has a phase-type distribution, that is, the distribution of the time the Markov process spends in non-absorbing states (or phases) before the absorbing state is reached (Neuts [50], Asmussen et al. [1]). We also give explicit approximations of the transition intensities of the Markov process. This includes transitions between non-adjacent mutations through stochastic tunneling (Carter and Wagner [11], Komarova et al. [40], Iwasa et al. [32]), where the intermediate genetic variants (the tunnel) are kept at a low frequency.

The paper is organized as follows: In Sect. 12.2 we introduce the framework for how the genetic composition of the population evolves over time by means of a Moran model (Moran [49], Section 3.4 of Ewens [21]) with deaths, births, and mutations with different selective fitness. In Sect. 12.3 we introduce the Markov process for fixed population states when the mutation rates are smaller than the inverse population size, and define the phase-type distribution for the time until this Markov process reaches its absorbing state. Then in Sect. 12.4 we give conditions under which the waiting time until the last mutant gets fixed, converges weakly towards a phase-type distribution, as the size of the population grows. After stating some results for the fixation of one single mutant in Sect. 12.5, we then provide explicit approximations, in Sect. 12.6, of the transition rates of the Markov process between different fixed states. Then we illustrate the theory for a number of asymptotic scenarios in Sect. 12.7, provide some adjustments of the asymptotic theory in Sect. 12.8, and give a summary with further extensions in Sect. 12.9. In Appendix A we provide a simulation algorithm, in Appendix B we derive an explicit approximation of the expected waiting time for one single mutant to get fixed, and in Appendix C we sketch proofs of main results.

12.2 Moran Model with Mutations and Selection

Consider a homogeneous and haploid population of constant size that consists of N individuals, all of which have the same sex. Each individual has one of $m+1$ possible types $0, 1, \ldots, m$. We can think of these types as different genetic variants or alleles, where 0 is a wildtype allele that is modified by m successive mutations. The genetic composition of the population is summarized by means of an $(m+1)$-dimensional vector

$$Z_t = (Z_{t0}, \ldots, Z_{tm}) \in \mathcal{Z}, \tag{12.1}$$

whose components represent the fraction of all alleles at time $t \geq 0$. It is assumed that t is a continuous parameter counted in units of generations. The allele frequency

configuration (12.1) is a stochastic process whose state space $\mathcal{Z}$ is the intersection of the m-simplex

$$\Delta = \left\{ z = (z_0, \ldots, z_m);\ z_i \geq 0, \sum_{i=0}^{m} z_i = 1 \right\}$$

spanned by the vectors

$$e_0 = (0, \ldots, 0), e_1 = (1, 0, \ldots, 0), \ldots, e_m = (0, 0, \ldots, 1),$$

and the set $\mathbb{N}^{(m+1)}/N$ of vectors z whose coordinates are natural numbers divided by N. More specifically, we will assume that (12.1) is a Moran model, where mutations between neighboring types $i \to i+1$ and $i \to i-1$ are possible, and where individuals with allele i have a selective fitness s_i, with $s_0 = 1$ and $s_i > 0$ for $i = 1, \ldots, m$. In our model, these numbers correspond to negative selection, neutral selection, and positive selection for allele i, depending on whether $0 < s_i < 1$, $s_i = 1$, and $s_i > 1$ respectively. The population starts with all individuals having type 0, so that $Z_0 = e_0$. It has overlapping generations, with a reproduction scheme defined as follows:

(i) Each individual dies independently according to a Poisson process with rate 1.
(ii) When an individuals dies, an offspring of some randomly chosen individual (including the one that dies) replaces it. The parent is chosen among the N individuals in the population, with probabilities proportional to their selection coefficients s_i.
(iii) If the parent has type $i < m$, the offspring in step (ii) mutates to $i+1$ with probability $u_{i+1} > 0$ and to $i-1$ with probability $v_{i-1} \geq 0$ (with $v_{-1} = 0$).

It follows from these reproduction rules that $\{Z_t;\ t \geq 0\}$ is a continuous time and time homogeneous Markov process on $\mathcal{Z}$. Our primary objects of study are the waiting time

$$T_m = \inf\{t \geq 0;\ Z_t = e_m\} \tag{12.2}$$

until allele m gets fixed in the population; and the waiting time

$$\tilde{T}_m = \inf\{t \geq 0;\ Z_{tm} > 0\} \tag{12.3}$$

until this allele first appears. Notice that T_m is the time until Z_t reaches the absorbing state e_m. On the other hand, $\tilde{T}_m$ is the hitting time of $\mathcal{Z}_m = \{z = (z_0, \ldots, z_m) \in \mathcal{Z};\ z_m > 0\}$, which is not an absorbing set of states, since the descendants of a type m individual may die out before this allele gets fixed in the population. However, if we modify the dynamics of Z_t and stop it as soon as it reaches $\mathcal{Z}_m$, we may treat this set as one single absorbing state.

It is also possible to allow for backward mutations when offspring of type m individuals are born, with probability v_{m-1}. Although this will affect the distribution of T_m, it will not impact the approximations of this distribution that we discuss in the following sections.

12.3 Phase-Type Distribution Approximation of Waiting Time

12.3.1 Asymptotic Notation

We will analyze the waiting times $T_m = T_{m,N}$ and $\tilde{T}_m = \tilde{T}_{m,N}$ asymptotically as the population size N tends to infinity. The various parameters of the model will in general depend on N as well, such as $u_i = u_{i,N}$, $v_i = v_{i,N}$, and $s_i = s_{i,N}$. We will use Bachmann–Landau asymptotic notation as $N \to \infty$, for instance $a_N \sim b_N$ if $a_N/b_N \to 1$, $b_N = O(a_N)$ if b_N/a_N stays bounded, $b_N = \Omega(a_N)$ if b_N/a_N is bounded away from zero, $b_N = \Theta(a_N)$ if a_N and b_N are of the same order (that is, $b_N = O(a_N)$ and $b_N = \Omega(a_N)$), and $b_N = o(a_N)$ if $b_N/a_N \to 0$. We will also make use of the analogous notation for sequences of random variables Y_N, with $Y_N = O_p(a_N)$ if Y_N/a_N stays bounded in probability and $Y_N = o_p(a_N)$ if Y_N/a_N converges to zero in probability. Suppose Y is a random variable with distribution F. We denote this as $Y \overset{\mathcal{L}}{\in} F$, and if Y_N is a sequence of random variables converging weakly towards Y, we often use the shorthand notation $Y_N \xrightarrow{\mathcal{L}} F$. For simplicity of notation, we will mostly omit index N for sequences of numbers or random variables that are functions of N. Sometimes, we also write $a_N \ll b_N$ or $b_N \gg a_N$ instead of $a_N = o(b_N)$.

12.3.2 Simplified Markov Process Between Fixed Population States

For all but very small N, it is not possible to get explicit and easily computable expressions for the distributions of T_m and $\tilde{T}_m$, since the state space $\mathcal{Z}$ gets huge when N grows. It is however possible to get accurate approximations of these distributions under appropriate conditions. The most crucial assumption is that the forward and backward mutation rates u_i and v_i tend to zero at a rate faster than the inverse population size, i.e.

$$u_i = o(N^{-1}), \tag{12.4}$$

and

$$v_i = o(N^{-1}) \tag{12.5}$$

for all i as $N \to \infty$. The implication of (12.4)–(12.5) is that most of the time, all individuals of the population will have the same type, and changes of this type occur rapidly when an individual with a new mutation gets many descendants that eventually take over the population. For this reason it may appear that a certain allele $i < m$ has been fixed permanently. But this is only temporary, since forward or backward mutations may later drive the population towards other fixed states.

This phenomenon was referred to as quasi-fixation in Hössjer et al. [31] for certain one-way mutation models with two possible alleles. Because of these rapid changes of the genetic decomposition Z_t of the population, it is well approximated by a continuous time Markov process defined on the finite subset

$$\mathcal{Z}_{\text{hom}} = \{e_0, \ldots, e_m\} \tag{12.6}$$

of $\mathcal{Z}$ that consists of all possible states of a type-homogeneous population. Here e_i refers to fixed state i of the population, so that all its individuals have the same type i. The simplified process has intensity matrix $\Lambda = (\lambda_{ij})_{i,j=0}^{m}$, where $\lambda_{ij} > 0$ is the rate of jumping from e_i to e_j when $j \neq i$, and $-\lambda_{ii} = \sum_{j;\, j\neq i} \lambda_{ij}$ is the rate of leaving e_i. Since e_m is an absorbing state, the intensity matrix can be decomposed as

$$\Lambda = \begin{pmatrix} \Lambda_0 & \lambda \\ 0 & 0 \end{pmatrix}, \tag{12.7}$$

where $\Lambda_0 = (\lambda_{i,j})_{i,j=0}^{m-1}$ contains the transition rates from and among the non-absorbing states, $0 = (0, \ldots, 0)$ is a row vector with m zeros, T denotes vector transposition, and $\lambda = (\lambda_{0m}, \ldots, \lambda_{m-1,m})^T$ is a column vector containing the transition rates from all non-absorbing states to e_m. A transition of Z_t from e_i to e_j corresponds to a stochastic tunneling event when $|j - i| \geq 2$. For instance, when $j \geq i + 2$, it represents a scenario where some individual who lives in a homogeneous type i population, has descendants from the same line of descent that experience mutations to $i + 1, i + 2, \ldots, j$, and then type j spreads to the whole population before any of the intermediate types do.

12.3.3 Defining Phase-Type Distribution Approximation

Since T_m is the time until the absorbing state e_m is reached, the simplified Markov process assumption with state space (12.6) and intensity matrix (12.7) implies that approximately

$$T_m \overset{\mathcal{L}}{\in} \text{PD}(\tilde{e}_0, \Lambda_0) \tag{12.8}$$

has a phase-type distribution, where $\tilde{e}_i$ is a unit vector of length m that contains the first m components of e_i. This phase-type distribution has two arguments, where the first, $\tilde{e}_0$, refers to the starting distribution of the Markov process among the non-absorbing states, and the second argument gives the intensity matrix among and from the non-absorbing states. From (12.8) we get very explicit approximate expressions for the density function

$$f_{T_m}(t) = \tilde{e}_0 \exp(\Lambda_0 t)\lambda, \quad t > 0, \tag{12.9}$$

the expected value

$$E(T_m) = -\tilde{e}_0 \Lambda_0^{-1} 1 \tag{12.10}$$

and the variance

$$\mathrm{Var}(T_m) = 2\tilde{e}_0 \Lambda_0^{-2} 1 - \left(\tilde{e}_0 \Lambda_0^{-1} 1\right)^2 \tag{12.11}$$

of T_m, with $1 = (1, \ldots, 1)^T$ a column vector of m ones.

In order to approximate the law of the waiting time $\tilde{T}_m$ in (12.3), we approximate $\{Z_t\}$ by a Markov process on

$$\tilde{\mathcal{Z}}_{\mathrm{hom}} = \{e_0, \ldots, e_{m-1}, \mathcal{Z}_m\}. \tag{12.12}$$

Then we may use (12.8) as a distributional approximation for $\tilde{T}_m$ rather than T_m, if λ_{im} is interpreted as a transition rate from e_i to $\mathcal{Z}_m$ when $i < m$, rather than from e_i to e_m.

12.4 Waiting Time Asymptotics

12.4.1 Regularity Conditions

In order to formulate precise asymptotic distributional results for T_m and $\tilde{T}_m$ when $N \to \infty$, we need some additional definitions and assumptions. We will focus on T_m, and then briefly point out the differences for $\tilde{T}_m$.

As a first step, let $\{\tau_k\}_{k=0}^M$ be the time points when a new allele gets fixed in the population. They are defined recursively as $\tau_0 = 0$ and

$$\tau_k = \inf\{t > \tau_{k-1};\ Z_t \in \{e_0, \ldots, e_m\} \setminus Z_{\tau_{k-1}}\}, \tag{12.13}$$

for $k = 1, 2, \ldots, M$, with $\tau_M = T_m$ the time point when Z_t reaches its absorbing state e_m. Clearly, $\{Z_{\tau_k};\ k = 0, 1, \ldots\}$ is a Markov chain with state space $\mathcal{Z}_{\mathrm{hom}}$ and transition probabilities

$$p_{ij} = p_{ij,N} = \begin{cases} P(Z_{\tau_{k+1}} = e_j | Z_{\tau_k} = e_i), & i = 0, \ldots, m-1, \\ & j = 0, \ldots, m, \\ 0, & i = m,\ j = 0, \ldots, m-1. \end{cases} \tag{12.14}$$

Since $Z_{\tau_k} \neq Z_{\tau_{k+1}}$ for $k < M$, the diagonal elements of the transition matrix $P = (p_{ij})_{i,j=0}^m$ vanish for all non-absorbing states, i.e. $p_{ii} = 0$ for $i < m$.

Assume that $Z_{\tau_k} = e_i$ for some $k < M$ and $i < m$. We will study what happens between the time points τ_k and τ_{k+1}, and refer to a forward mutation $i \to i + 1$ as successful if its descendants eventually take over the population. Let $f_i = f_{i,N}$ be the probability that a forward mutation that happens while all individuals of the

population have the same type i, is successful. Likewise, if $i > 0$, a successful backward mutation of type $i \to i-1$ is one whose descendants eventually take over the population. Denote by $b_i = b_{i,N}$ the probability that a backward mutation is successful, given that it happens in a homogeneous type i population. For definiteness, we also put $b_0 = 0$. A successful mutation from type i is either forward or backward, and due to (12.4)–(12.5), it will arrive when the population is homogeneous or almost homogeneous of type i. Therefore, a successful mutation from type i arrives at a rate close to

$$\mu_i = \mu_{i,N} = N v_{i-1} b_i + N u_{i+1} f_i, \quad i = 0, \ldots, m-1, \tag{12.15}$$

since new backward and forward mutations appear at rates $N v_{i-1}$ and $N u_{i+1}$ among N individuals with the same type i, but only a fraction b_i and f_i of them are successful, and cause a change in the population to another fixed state. For the absorbing state we put $\mu_m = 0$.

There will be at least one successful mutation within (τ_k, τ_{k+1}), and let τ'_{k+1} be the time point when the first of these mutations arrives. We will assume below that $\tau_{k+1} - \tau'_{k+1}$ is asymptotically negligible in comparison to the total waiting time T_m, which reflects that fact that all transitions of Z_t occur rapidly. This suggests that it is asymptotically accurate to use transition rates

$$\lambda_{ij} = \begin{cases} -\mu_i, & i = j, \\ \mu_i p_{ij}, & i \neq j, \end{cases} \tag{12.16}$$

in (12.8). In order to verify this we need to make some additional assumptions on how the rates in (12.16) behave as $N \to \infty$. We will first of all assume that the transition probabilities in (12.14) satisfy

$$p_{ij} \to \pi_{ij}, \quad i, j = 0, 1, \ldots, m, \tag{12.17}$$

as $N \to \infty$, so that $\Pi = (\pi_{ij})_{i,j=0}^{m}$ is the asymptotic transition matrix of the embedded Markov chain $\{Z_{\tau_k};\ k = 0, 1, \ldots\}$. We will then postulate that

$$(I - \Pi_0)^{-1} \text{ is invertible}, \tag{12.18}$$

with I the identity matrix of order m and Π_0 a square matrix of order m that contains the first m rows and first m columns of Π. Condition (12.18) guarantees that the asymptotic Markov chain reaches its absorbing state e_m with probability 1, since it implies $P(M < \infty) = \tilde{e}_0 (I - \Pi_0)^{-1} \pi = 1$, where $\pi = (\pi_{0m}, \ldots, \pi_{m-1,m})^T$. Let

$$I_{\text{as}} = \{i;\ 0 \leq i \leq m-1, \tilde{e}_0 (I - \Pi_0)^{-1} \tilde{e}_i^T > 0\} \tag{12.19}$$

refer to the asymptotic states. It consists of those non-absorbing states that are visited with a positive probability asymptotically as $N \to \infty$, since $i \in I_{\text{as}}$ is

equivalent to requiring that $(\Pi^k)_{0i} > 0$ for at least one $k = 0, 1, \ldots$. The remaining non-asymptotic states are denoted as

$$I_{\text{nas}} = \{1, \ldots, m-1\} \setminus I_{\text{as}}. \tag{12.20}$$

Among the asymptotic states, it is also important to know for how long time they are visited. We therefore express the expected waiting time

$$E(T_m) = E_0 + E_1 + \cdots + E_{m-1}$$

until allele m gets fixed as a sum of m terms, with $E_i = E_{i,N} = -\tilde{e}_0 \Lambda_0^{-1} \tilde{e}_i^T$ the expected time spent in state e_i before absorption into state e_m takes place. Notice that Λ_0 is invertible for each finite N, and therefore each E_i is well defined with $0 < E_i < \infty$. Indeed, since all $u_i > 0$, it follows that any fixed population state e_j with $j = i+1, \ldots, m$ can be reached from fixed population state e_i in one step. Therefore, all entries of Λ_0 above the diagonal are strictly positive, whereas the diagonal elements and row sums of Λ_0 are strictly negative. From the Gershgorin Circle Theorem we deduce that all eigenvalues of Λ_0 have a strictly negative real part, so that Λ_0 is invertible. We will assume that the limits

$$E_i / E(T_m) \to c_i, \quad 0, 1, \ldots, m-1, \tag{12.21}$$

exist as $N \to \infty$, and define

$$I_{\text{long}} = \{i;\ 0 \le i \le m-1, c_i > 0\} \tag{12.22}$$

as the set of asymptotic states e_i that are visited for such a long time that they have an asymptotic contribution to the expected waiting time (12.10). We also put

$$I_{\text{short}} = I_{\text{as}} \setminus I_{\text{long}} \tag{12.23}$$

for those non-absorbing states that are asymptotic, but visits to them are too short to have an asymptotic impact on the expected waiting time. It follows from (12.21) that the transition rates from the states in I_{long} have the same order

$$\mu_{\min} = \mu_{\min,N} = \min\{\mu_i;\ i \in I_{\text{long}}\}, \tag{12.24}$$

and it is the inverse of (12.24) that determines the asymptotic size of the waiting time (12.2). We will therefore rescale time in units of $\mu_{\min}$ and assume that

$$\frac{\mu_i}{\mu_{\min}} \to \kappa_i, \quad i = 0, \ldots, m, \tag{12.25}$$

as $N \to \infty$, where the normalized rate κ_i of leaving state e_i satisfies $1 \le \kappa_i < \infty$ for $i \in I_{\text{long}}$, $\kappa_i = \infty$ for $i \in I_{\text{short}}$, $0 \le \kappa_i \le \infty$ for $i \in I_{\text{nas}}$, and $\kappa_m = 0$. In order to

ensure that the time between the appearance of a successful mutation and fixation of a new allele is asymptotically negligible, we assume that

$$P\left(\tau_{k+1} - \tau'_{k+1} > \varepsilon \mu_{\min}^{-1} | Z_{\tau_k} = e_i\right) \to 1 \ \forall \varepsilon > 0 \text{ and } i \in I_{\text{as}}, \tag{12.26}$$

as $N \to \infty$. Notice that the probability on the left hand side of (12.26) does not depend on k, because of the Markov property of $\{Z_{\tau_k}\}$.

12.4.2 Main Results on Waiting Time Asymptotics

The following theorem specifies the asymptotic phase-type distribution of the waiting time T_m until allele m gets fixed. A proof of it is sketched in Appendix C.

Theorem 12.1 *Consider a Moran model for a population of size N with types (alleles) $0, \ldots, m$ that starts with all its individuals in allelic state 0 and then reproduces according to (i)–(iii) of Sect. 12.2, so that forward ($i \to i+1$) and backward ($i \to i-1$) mutations between nearby alleles are possible. Assume that the forward and backward mutation rates satisfy (12.4)–(12.5), that the transition probabilities between fixed population states where all individuals have the same allele, converge as in (12.17)–(12.18), that the expected times spent in various fixed states converge as in (12.21), that the rates of leaving the various fixed states satisfy (12.25), and that the time between appearance of a new successful mutation and fixation of a new allele is asymptotically negligible (12.26), as $N \to \infty$. Then the waiting time T_m until allele m gets fixed has a phase-type distribution*

$$\mu_{min} T_m \xrightarrow{\mathcal{L}} PD(\tilde{e}_0, \Sigma_0), \tag{12.27}$$

asymptotically as $N \to \infty$, when rescaled by μ_{min} in (12.24), the minimal rate of leaving a fixed state, among those that are visited for a positive fraction of time. The second argument Σ_0 on the right hand side of (12.27) contains the first m rows and first m columns of the intensity matrix $\Sigma = (\Sigma_{ij})_{i,j=0}^{m}$, with

$$\Sigma_{ij} = \begin{cases} -\kappa_i, & j = i, \\ \kappa_i \pi_{ij}, & j \neq i, \end{cases} \tag{12.28}$$

π_{ij} is the asymptotic transition probability (12.17) between fixed states i and j, and κ_i is the normalized rate (12.25) of leaving fixed state i.

We will make some comments on the asymptotic inverse size $\mu_{\min}$ of the waiting time T_m, and the matrix Σ_0 of the limit distribution in (12.27). In some applications, it convenient to generalize (12.24) and let

$$\mu_{\min} = C \min\{\mu_i;\ i \in I_{\text{long}}\} \tag{12.29}$$

for some constant $C > 0$, chosen in order to get a simple expression for $\mu_{\min}$. It is straightforward to see that Theorem 12.1 remains unchanged with this minor modification. The matrix Σ_0 contains asymptotic transition rates among and from all non-absorbing states, after the change of time scale in (12.24). It will be degenerate when either I_{short} or I_{nas} are non-empty. However, it turns out that (12.27) is still well defined, if we disregard those rows and columns of Σ_0 that correspond to I_{nas} and take the limit $\kappa_i \to \infty$ for all $i \in I_{\text{short}}$.

Consider the special case when either all $v_i = 0$, or that backward mutations have no asymptotic impact on the waiting time distribution. An important instance of (12.27) occurs if, in addition, successful forward mutations in a homogeneous type i environment always causes the same allele $F(i) > i$ to get fixed in the population, i.e.

$$\pi_{i,F(i)} = 1, \quad i \in I_{\text{as}}. \tag{12.30}$$

With this extra regularity condition, we obtain the following corollary of Theorem 12.1:

Corollary 12.1 *Consider the Moran model of Sect. 12.2. Assume that the conditions of Theorem 12.1 hold, that only forward mutations have an asymptotic impact on the population dynamics in such a way that the forward jumps between fixed population states occur according to* (12.30)*. Then the waiting time T_m until allele m gets fixed has a hypoexponential limit distribution*

$$\mu_{min} T_m \xrightarrow{\mathcal{L}} \sum_{i \in I_{long}} \kappa_i^{-1} X_i \tag{12.31}$$

as $N \to \infty$, where $X_0, \ldots, X_{m-1}$ are independent and exponentially distributed random variables with expected value 1.

Remark 12.1 The asymptotic result for the waiting time distribution of $\tilde{T}_m$ is analogous to (12.27), if we replace e_m by $\mathcal{Z}_m$ in all definitions. In particular, we interpret Σ_{im} as a normalized transition rates from e_i to $\mathcal{Z}_m$ (rather than to e_m) for $i = 0, \ldots, m-1$.

12.5 Fixation in a Two Type Moran Model Without Mutations

As a preparation for the next sections, we will state two well known result on the fixation probability and expected time to fixation, for a Moran model with two alleles ($m = 1$) and no mutations. If these two alleles start at frequencies $N - 1$ and 1, and have selection coefficients 1 and $s > 0$ respectively,

$$\beta(s) = \beta_N(s) = \begin{cases} 1/N, & s = 1, \\ (1 - s^{-1})/(1 - s^{-N}), & s \neq 1 \end{cases} \tag{12.32}$$

is the probability that the second allele gets fixed, whereas $1-\beta(s)$ is the probability that the first allele does (Komarova et al. [40], Section 6.1 of Durrett [17]). We will make frequent use of asymptotic expressions for the fixation probability of large populations. It follows from (12.32) that

$$\beta(s) \sim \begin{cases} (s^{-1}-1)\times s^N, & 1-s \gg 1/N, \\ x/\left[1-\exp(-x)\right]\times 1/N, & s=1+x/N, \\ 1-s^{-1}, & s-1 \gg 1/N, \end{cases} \tag{12.33}$$

as $N\to\infty$, where $x\neq 0$ in the second line is a constant, not depending on N.

Given that the second alleles takes over, we let $\alpha(s)$ be the expected time it takes for this to happen. Kimura and Ohta [38] derived a general diffusion approximation of $\alpha(s)$ for a large class of models with two alleles, see also Section 8.9 of Crow and Kimura [15], or Theorems 1.32 and 6.3 of Durrett [17]. In Appendix B we calculate this diffusion approximation $\alpha(s)$ for the Moran model of Sect. 12.2. In particular, we show that this diffusion approximation is of the order

$$\alpha(s)=\alpha_N(s)\sim \begin{cases} (1+s)\log(N)/(1-s), & \text{if } s<1, \\ N, & \text{if } s=1, \\ (1+s)\log(N)/(s-1), & \text{if } s>1, \end{cases} \tag{12.34}$$

asymptotically as $N\to\infty$, if s is kept fixed. The expected time to fixation in (12.34) is much different for neutral and non-neutral alleles. This is also true for the more accurate diffusion approximation of $\alpha(s)$ in Appendix B, although it has a somewhat smoother transition between $s=1$ and $s\neq 1$.

12.6 Explicit Approximate Transition Rates Between Fixed Population States

Returning to the general model with m mutations, we recall that Theorem 12.1 gives quite general conditions under which the normalized waiting times $\mu_{\min}T_m$ and $\mu_{\min}\tilde{T}_m$ have asymptotic phase-type distributions as the population size $N\to\infty$. Under these assumptions the unnormalized waiting times T_m (cf. (12.8)) and $\tilde{T}_m$ are also well approximated by phase-type distributions. But in order to apply these results we still need to find explicit approximations of the Markov transition rates λ_{ij} in (12.8) and (12.15)–(12.16) between fixed population states. As in Sect. 12.4 we focus on T_m and then pinpoint the difference when $\tilde{T}_m$ is of interest.

12.6.1 Defining Approximate Transition Rates

We introduce

$$\hat{\lambda}_{ij} = \begin{cases} N u_{i+1} r_{ij} \beta(s_j/s_i), & j > i, \\ N v_{i-1} r_{ij} \beta(s_j/s_i), & j < i \end{cases} \tag{12.35}$$

as an approximation of λ_{ij} when $i < m$ and $j \neq i$. The quantity $r_{ij} = \hat{q}_{ij}$ approximates a certain probability q_{ij}. When $|j - i| = 1$ we put $q_{i,i-1} = q_{i,i+1} = 1$. When $j \geq i + 2$, q_{ij} is a probability of tunneling from $i + 1$ to j. In more detail, q_{ij} is the probability that a forward mutation $i \to i + 1$, that occurs in a homogeneous type i population, gets a at least one descendant that mutates from $j - 1$ to j before any other allele gets fixed. Analogously when $j \leq i - 2$, q_{ij} is the probability of tunneling backwards from $i - 1$ to j. That is, q_{ij} is the probability for a backward mutation $i \to i - 1$, that occurs in a homogeneous population of type i individuals, to get at least one descendant that mutates to from $j + 1$ to j before any other allele gets fixed. For definiteness, we also put $\hat{\lambda}_{mj} = 0$ for all j.

It follows from (12.32) that the $\beta(s_j/s_i)$ term of (12.35) is the probability that descendants of one single type j individual take over a population where all the others have type i, if further mutations do not occur. In our setting, it is an approximation of the probability that the descendants of the individual that first mutated into j, take over the population before any new mutations occur. In order for this approximation to be accurate, it is required that no other allele than i attains a high frequency before j gets fixed (recall that the type j mutation itself was a descendant of a successful $i \to i \pm 1$ mutation, that appeared in homogeneous or almost homogeneous type i population).

In order to finalize the definition of $\hat{\lambda}_{ij}$ in (12.35) we must specify how r_{ij} approximates q_{ij}. When $|j - i| = 1$ we put $r_{ij} = 1$. When $|i - j| \geq 2$, we introduce explicit approximations

$$r_{ij} = \begin{cases} \prod_{l=i+1}^{j-1} R(\rho_{ilj})^{2^{-(l-i-1)}} u_{l+1}^{2^{-(l-i)}}, & j > i, \\ \prod_{l=j+1}^{i-1} R(\rho_{ilj})^{2^{-(i-l-1)}} v_{l-1}^{2^{-(i-l)}}, & j < i \end{cases} \tag{12.36}$$

of the tunneling probabilities q_{ij} that are accurate when no other allele reaches a high frequency during a transition from i to j. The parameters ρ_{ilj} in (12.36) quantify the difference between selection coefficients s_i and s_l, on a scale determined by a tunneling probability from l to j. When $i < j$, they are defined through

$$\frac{s_l}{s_i} = 1 + \rho_{ilj} \sqrt{u_{l+1} r_{ilj}} \tag{12.37}$$

for $l = j - 1, \ldots, i + 1$. This involves some other quantities r_{ilj}, which are also defined recursively, for $l = j - 1, \ldots, i$, starting with $r_{i,j-1,j} = 1$ and then using the relation

$$r_{ilj} = R(\rho_{i,l+1,j}) \sqrt{r_{i,l+1,j} u_{l+2}}. \tag{12.38}$$

When this recursion has stopped at $l = i$ we obtain the upper row of (12.36) by putting $r_{ij} = r_{iij}$. Here r_{ilj} approximates the probability q_{ilj} that a mutation $l \to l+1$, which occurs in a homogeneous type i population, gets a least one descendant that mutates into j, before any other allele gets fixed. In particular, $q_{iij} = q_{ij}$. Similarly, when $i > j$, we have that

$$\frac{s_l}{s_i} = 1 + \rho_{ilj}\sqrt{\nu_{l-1} r_{ilj}} \tag{12.39}$$

for $l = j+1, \ldots, i-1$. The probabilities r_{ilj} are defined recursively for $l = j+1, \ldots, i$, starting with $r_{i,j+1,j} = 1$, then iterating

$$r_{ilj} = R(\rho_{i,l-1,j})\sqrt{r_{i,l-1,j}\nu_{l-2}}, \tag{12.40}$$

and finally getting the lower row of (12.36) from $r_{ij} = r_{iij}$. The function

$$R(\rho) = \frac{\sqrt{\rho^2+4}+\rho}{2} \tag{12.41}$$

specifies the way in which differences between the selection coefficients $s_i, \ldots, s_{j-1}$ affect the probability r_{ij} in (12.36), see also equation (10) of Iwasa et al. [32] or equation (5) of Durrett and Schmidt [19]. Intuitively, if type l is more fit than i, then $\rho_{ilj} > 0$, and the probability in (12.36) increases (since $R(\rho) > 1$ when $\rho > 0$, in particular $R(\rho) \sim \rho$ when $\rho \to \infty$), if $s_l = s_i$ then $\rho_{ilj} = 0$ will have no impact on r_{ij} (since $R(0) = 1$), and finally, if l is less fit than i and therefore $\rho_{ilj} < 0$, this will decrease the probability in (12.36) (since $R(\rho) < 1$ when $\rho < 0$, in particular $R(\rho) \sim -1/\rho$ as $\rho \to -\infty$).

It is possible to obtain a more accurate approximation of q_{ij} than (12.36), without using the quantities ρ_{ilj} nor the function R (see the end of the proof of Lemma 12.3 in Appendix C for details). Formula (12.36) is more explicit though, and therefore it gives more insight into how the mutation rates and the selection coefficients affect the approximate tunneling probabilities r_{ij}.

12.6.2 Conditions Under Which Approximate Transition Rates Are Accurate

It turns out that Eqs. (12.35) and (12.36) are good approximations of λ_{ij} for those forward transition rates ($j > i$) and backward transition rates ($j < i$) that dominate asymptotically, provided there is exactly one forward rate and at most one backward rate from i that dominate. This can be formulated as follows. Define

$$\hat{\mu}_i = -\hat{\lambda}_{ii} = \sum_{j;j\neq i} \hat{\lambda}_{ij} \text{ for } i = 0, \ldots, m-1, \tag{12.42}$$

as an approximation of the rate μ_i in (12.15) at which a successful mutation occurs in a type i population, and suppose that

$$\hat{p}_{ij} = \frac{\hat{\lambda}_{ij}}{\hat{\mu}_i} \to \hat{\pi}_{ij} \tag{12.43}$$

as $N \to \infty$ for all $0 \le i \le m-1$ and $j \ne i$. For definiteness we also put $\hat{\pi}_{ii} = 0$ and $\hat{\pi}_{mj} = 0$ for $j = 0, \ldots, m-1$. We assume there is at most one index $0 \le B(i) < i$ for each $i = 1, \ldots, m-1$, and exactly one index $i < F(i) \le m$ for each $i = 0, \ldots, m-1$, such that fixation events from i will always be to $B(i)$ for backward mutations, and to $F(i)$ for forward mutations. This can phrased as

$$\begin{aligned} \hat{\pi}_{0,F(0)} &= 1, \\ \hat{\pi}_{i,B(i)} + \hat{\pi}_{i,F(i)} &= 1, i = 0, \ldots, m-1, \\ \hat{\pi}_{i,F(i)} &> 0, i = 1, \ldots, m-1, \end{aligned} \tag{12.44}$$

with $\hat{\pi}_{i,B(i)} = 0$ in the middle equation when $B(i) = \emptyset$, i.e. when backward mutations from a type i population have no asymptotic impact. In particular, the forward fixation from i involves stochastic tunneling if $F(i) \ge i+2$. In order for this to happen, $\hat{\lambda}_{iF(i)}$ must have a larger order asymptotically than all other $\hat{\lambda}_{ij}$ with $j > i$. It follows from (12.36) and some of the regularity conditions below, that a necessary condition for this to happen is that type j is more beneficial for reproduction than all the intermediate alleles. A similar condition applies for backward mutations, and we can summarize these necessary tunneling conditions as follows:

$$\begin{aligned} F(i) \ge i+2 &\Longrightarrow s_{F(i)} > \max\left(s_{i+1}, \ldots, s_{F(i)-1}\right), \\ B(i) \le i-2 &\Longrightarrow s_{B(i)} > \max\left(s_{B(i)+1}, \ldots, s_{i-1}\right), \end{aligned} \tag{12.45}$$

where the lower equation only applies when $B(i) \ne \emptyset$. We will need some additional regularity conditions. The first one consists of four relations

$$\begin{aligned} u_i/u_{i+1} &= O(1), && i = 0, \ldots, m-1, \\ v_{i-1} &= O(u_{i+1}), && i = 1, \ldots, m-1, \\ v_j/v_{j-1} &= O(1), && \text{if } B(i) < j < i \text{ for some } i = 2, \ldots, m-1, \\ u_{j+1} &= O(v_{j-1}), && \text{if } B(i) < j < i \text{ for some } i = 2, \ldots, m-1, \end{aligned} \tag{12.46}$$

each of which imposes some restrictions on the mutation rates. The second and fourth relations of (12.46) guarantee that backward mutations will have no asymptotic impact on forward fixations, and vice versa. The first equation of (12.46) requires that mutation rates to higher types are at least of the same order as mutation rates to lower types. Otherwise forward stochastic tunneling will be more difficult, and the formulas for some of the tunneling probabilities in (12.36) will look different. The third equation of (12.46) is the analogous requirement on backward mutations.

Notice that neither the third nor the fourth relation of (12.46) apply when back mutations do not exist or have no asymptotic impact, i.e. if $B(i) = \emptyset$ for all i.

In order to assure that condition (12.26) holds, i.e. that the time for successful mutations to get fixed are asymptotically negligible, we will assume that

$$\begin{aligned} &N \min_{i \in I_{\text{long}}} u_{F(i)}^{2-2^{-(F(i)-i-1)}} \beta(s_{F(i)}/s_i) \\ &= o\left(\min_{i \in I_{\text{as}}} \min\left[\alpha^{-1}\left(\tfrac{s_{B(i)}}{s_i}\right), \alpha^{-1}\left(\tfrac{s_{F(i)}}{s_i}\right)\right]\right), \end{aligned} \tag{12.47}$$

where $\beta(s)$ and $\alpha(s)$ are the fixation probability (12.32) and expected fixation time (12.34), respectively. If (12.47) does not hold, T_m will not only be affected by the waiting times for successful mutations to occur, but their fixation time will also have an impact.

The next regularity condition requires that the parameters of Eqs. (12.37)–(12.39) are bounded, i.e.

$$|\rho_{ilj}| = O(1), \quad i < l < j \text{ or } j < l < i \tag{12.48}$$

when $N \to \infty$. This means that the fitness $s_1, \ldots, s_{m-1}$ of the first $m-1$ mutant alleles approach 1 as N grows, so that each one of them is either slightly deleterious, neutral or slightly advantageous compared to the wildtype allele 0. The case of strong negative or positive selection ($\rho_{ilj} \to \pm\infty$ respectively) is not included in (12.48), but has been studied by Komarova et al. [40].

12.6.3 Asymptotic Distribution of Wating Time Based on Approximate Transition Rates

Equipped with the definitions and regularity conditions of Sect. 12.6.2, we are ready to formulate an asymptotic distributional result for the waiting time T_m (see Appendix C for a sketch of proof), where its limiting phase-type distribution can be derived from the explicit approximation (12.35) of the transition intensities between the fixed population states of the simplified Markov process.

Theorem 12.2 *Consider a Moran model for a population with N individuals and alleles $0, \ldots, m$ that starts with all its individuals in allelic state 0, and then reproduces according to (i)–(iii) of Sect. 12.2. Assume, as in Theorem 12.1, that (12.4)–(12.5), (12.17), (12.21), and (12.25) hold, and let λ_{ij} be the Markov transition rate (12.16) between two fixed population states i and j where all individuals have the same allele i and j respectively. Define $\hat{\lambda}_{ij}$ in (12.35) as an approximation of λ_{ij}, with $\hat{\mu}_i$ the approximate rate (12.42) of leaving state i and $\hat{\pi}_{ij}$ an approximation (12.43) of the probability π_{ij} in (12.17) of jumping from fixed state i to fixed state j, whereas $\hat{\mu}_{min} = \min_{i \in I_{long}} \hat{\mu}_i$ is an approximation of the minimal rate of leaving a fixed state, among those that are visited for a positive fraction of time. Assume further that (12.44)–(12.48) hold. Then $\pi_{ij} = \hat{\pi}_{ij}$ and $\hat{\mu}_i / \hat{\mu}_{min} \to \kappa_i$ for $i = 0, \ldots, m$*

as $N \to \infty$, where κ_i is the normalized rate (12.25) of leaving fixed state i. Moreover, the waiting time T_m until allele m gets fixed has an asymptotic phase-type distribution

$$\hat{\mu}_{min} T_m \xrightarrow{\mathcal{L}} PD(\tilde{e}_0, \Sigma_0) \tag{12.49}$$

as $N \to \infty$, where Σ_0 contains the first m rows and m columns of the intensity matrix Σ in (12.28).

Remark 12.2 The limit result for $\tilde{T}_m$ is analogous to Theorem 12.2. One simply puts $s_m = \infty$ everywhere, which corresponds to immediate fixation of a type m mutation, once it appears.

12.7 Illustrating the Theory

In this section we will illustrate Theorem 12.2. Recall that it gives the asymptotic waiting time distribution until the m:th mutant gets fixed in the population, based on the transition rates $\hat{\lambda}_{ij}$ in (12.35)–(12.36) that approximate λ_{ij} in (12.15)–(12.16). In order to determine the approximate waiting time distribution, it suffices to specify $\hat{\lambda}_{ij}$ for $i = 0, \ldots, m-1$ and $j \neq i$, $j = 0, \ldots, m$, and then look at the properties of these rates as the population size grows. We will consider different scenarios, not all of which satisfy the regularity conditions of Theorem 12.2. But in these cases we will argue why (12.49) still provides a fairly accurate asymptotic approximation of the waiting distribution T_m. On the other hand, it is implicit that the mutation rates are smaller than the inverse population size, according to (12.4)–(12.5), for all examples of this section.

12.7.1 The Case of Two Coordinated Mutations

When there are $m = 2$ coordinated mutations, formula (12.35) simplifies to

$$\begin{gathered}\hat{\lambda}_{01} = N u_1 \beta(s_1), \quad \hat{\lambda}_{02} = N R(\rho) u_1 u_2^{1/2} \beta(s_2), \quad \hat{\lambda}_{10} = N v \beta(1/s_1), \\ \hat{\lambda}_{12} = N u_2 \beta(s_2/s_1),\end{gathered} \tag{12.50}$$

where $v = v_0$, and $\rho = \rho_{012}$ is a real-valued constant (not depending on N) defined in (12.37) that is either negative, zero or positive. Here, this equation simplifies to

$$s_1 = 1 + \rho u_2^{1/2}. \tag{12.51}$$

We will investigate the limit distribution of the waiting time T_2 for different asymptotic scenarios.

12.7.1.1 No Backward Mutations, and Final Allele Has High Fitness

In this subsection we will make two favorable assumptions for the waiting time T_2; that there are no backward mutations ($v = 0$), and that allele 2 has a high fitness ($s_2 = \infty$). It turns out that the relative size of the two forward mutation rates, u_1 and u_2, is crucial for the asymptotic properties of T_2. We will look at four different cases.

Case 1: Second mutation rate very small. Assume that

$$u_2 = o(u_1 N^{-1}) \tag{12.52}$$

as $N \to \infty$. In this case, the three nonzero rates in (12.50) simplify to

$$\hat{\lambda}_{01} \sim u_1, \quad \hat{\lambda}_{02} = NR(\rho)u_1 u_2^{1/2}, \quad \hat{\lambda}_{12} = Nu_2. \tag{12.53}$$

It follows from (12.4) to (12.52) that $\hat{\lambda}_{02} \ll \hat{\lambda}_{01}$ and $\hat{\lambda}_{12} \ll \hat{\lambda}_{01}$, so that $\hat{\mu}_0 \sim \hat{\lambda}_{01} \gg \hat{\mu}_1 = Nu_2$. The asymptotic states with short and long waiting times are $I_{\text{short}} = \{0\}$ and $I_{\text{long}} = \{1\}$ respectively, the time rate to absorption is $\hat{\mu}_{\min} = Nu_2$, the mutation rates on the new time scale are $\kappa_0 = \infty$ and $\kappa_1 = 1$, and the nonzero asymptotic transition probabilities from the non-asymptotic states, are $\pi_{01} = \pi_{12} = 1$. This gives a normalized intensity matrix

$$\Sigma = \begin{pmatrix} -\infty & \infty & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

in (12.28). Because of the smallness of the second mutation rate u_2, there is asymptotically no tunneling from 0 to 2, but allele 1 gets fixed at first. After that it takes much longer time for the first allele 2 to arrive, in spite of the fact that this $1 \to 2$ mutation is successful with probability 1 (since $s_2 = \infty$, and therefore $\beta(s_2) = 1$). Consequently, the asymptotic distribution of T_2 will be dominated by the waiting time for allele 2 to appear, after allele 1 has first been fixed in the population, i.e.

$$Nu_2 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.54}$$

as $N \to \infty$. Notice that the exponential limit distribution in (12.54) is a special case of Corollary 12.1, although (12.52) violates regularity condition (12.46) of Theorem 12.2. However, this condition is only needed in order to get a good approximation of the tunneling rate $\hat{\lambda}_{02}$. But since stochastic tunneling $0 \to 2$ has no asymptotic impact ($\hat{\lambda}_{02} \ll \hat{\lambda}_{01}$), we still believe (12.54) is accurate.

Case 2: Second mutation rate small. If

$$u_1 N^{-1} \ll u_2 \ll N^{-2} \tag{12.55}$$

as $N \to \infty$, we get slightly different asymptotics compared to Case 1. The transition rates $\hat{\lambda}_{ij}$ are the same as in (12.53), but their asymptotic ordering $\hat{\lambda}_{02} \ll \hat{\lambda}_{01} \ll \hat{\lambda}_{12}$ is different. The states with short and long waiting time are therefore switched compared to Case 1 ($I_{\text{long}} = \{0\}$, $I_{\text{short}} = \{1\}$), with a time rate $\hat{\mu}_{\min} = \hat{\mu}_1 \sim \hat{\lambda}_{01} \sim u_1$ to absorption. The rescaled mutation rates from states 0 and 1 are $\kappa_0 = 1$ and $\kappa_1 = \infty$ respectively, whereas the nonzero asymptotic transition probabilities from the non-asymptotic states are the same as for Case 1 ($\pi_{01} = \pi_{12} = 1$). This gives a normalized intensity matrix

$$\Sigma = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -\infty & \infty \\ 0 & 0 & 0 \end{pmatrix}.$$

The second mutation rate u_2 in (12.55) is too small to allow for tunneling, but large enough to make the waiting time for allele 2 much shorter than the waiting time until allele 1 gets fixed at first. Notice that (12.55) allows for any of u_1 or u_2 to dominate asymptotically. In either case, the waiting time for allele 2 to fix is shorter, because of the selective advantage of this allele ($s_2 = \infty$). Therefore, the asymptotic distribution of T_2 will be dominated by the waiting time for allele 1 to fix, i.e.

$$u_1 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.56}$$

as $N \to \infty$. This limit result is also special case of Corollary 12.1, and it agrees with Theorem 2 of Durrett and Schmidt [19].

Case 3: Second mutation rate of intermediate size. We assume that

$$u_2 = \frac{\gamma}{N^2} \tag{12.57}$$

for some constant γ as $N \to \infty$. The transition rates in (12.50) then simplify to

$$\hat{\lambda}_{01} \sim u_1 \eta(\rho \gamma^{1/2}), \quad \hat{\lambda}_{02} = R(\rho) \gamma^{1/2} u_1, \quad \hat{\lambda}_{12} = \gamma / N, \tag{12.58}$$

where $\eta(x)$ is an asymptotic approximation of $N\beta(1 + x/N)$. From formula (12.33) we deduce that

$$\eta(x) = \begin{cases} 1, & x = 0, \\ x/\left[1 - \exp(-x)\right], & x \neq 0. \end{cases}$$

It follows from (12.4) and (12.58) that $\hat{\lambda}_{0i} \ll \hat{\lambda}_{12}$ for $i = 1, 2$, whereas $\hat{\lambda}_{01}$ and $\hat{\lambda}_{02}$ have the same asymptotic order. Therefore, $I_{\text{long}} = \{0\}$, $I_{\text{short}} = \{1\}$, $\hat{\mu}_{\min} = \hat{\mu}_0 = \hat{\lambda}_{01} + \hat{\lambda}_{02} \ll \hat{\mu}_1 = \hat{\lambda}_{12}$, $\kappa_0 = 1$, and $\kappa_1 = \infty$. This gives an asymptotic rescaled intensity matrix of the form

$$\Sigma = \begin{pmatrix} -1 & \pi_{01} & \pi_{02} \\ 0 & -\infty & \infty \\ 0 & 0 & 0 \end{pmatrix}, \tag{12.59}$$

where $\pi_{02} = \frac{R(\rho)\gamma^{1/2}}{R(\rho)\gamma^{1/2}+\eta(\rho\gamma^{1/2})}$ is the asymptotic probability for tunneling to occur, and $\pi_{01} = 1 - \pi_{02}$ is the corresponding probability of no tunneling. Since the two transition rates from allele 0 are of similar size asymptotically, allele 2 will either get fixed directly through stochastic tunneling, or in two steps where allele 1 spreads in the population at first, and then almost immediately after that, allele 2 takes over. Formula (12.49) suggests that

$$\left[\eta(\rho\gamma^{1/2}) + R(\rho)\gamma^{1/2}\right] u_1 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.60}$$

as $N \to \infty$. However, (12.60) is not correct since (12.44) is violated, that is, there is asymptotic competition between the two forward rates from allele 0, so that $F(0)$ does not exist. In order to see that (12.60) is wrong, consider the case when the intermediate allele 1 has the same selective advantage as allele 0 ($\rho = 0$). Then (12.60) simplifies to

$$(1 + \gamma^{1/2})u_1 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.61}$$

as $N \to \infty$, since $\eta(0) = R(0) = 1$. But this is different from Theorem 3 of Durrett et al. [20], which states that

$$\chi(\gamma)u_1 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.62}$$

as $N \to \infty$, where

$$\chi(\gamma) = \frac{\sum_{k=1}^{\infty} \frac{\gamma^k}{(k-1)!(k-1)!}}{\sum_{k=1}^{\infty} \frac{\gamma^k}{k!(k-1)!}}. \tag{12.63}$$

In order to quantify the difference between (12.61) and (12.62), we have plotted the ratio

$$\xi(\gamma) = \frac{1 + \gamma^{1/2}}{\chi(\gamma)} \tag{12.64}$$

of the two intensities in Fig. 12.1. It can be seen that the approximate intensity is always a bit larger than the exact one, with a maximum difference of 40%, although for most values of γ the difference is less than 20%. This implies that the approximate approach will underestimate the expected waiting time by up to 40%, since competition between the two fixation rates $0 \to 1$ and $0 \to 2$ is ignored. In Sect. 12.8 we will discuss a method that to some extent corrects for this.

Case 4: Second mutation rate large. Suppose

$$u_2 \gg N^{-2}, \tag{12.65}$$

so that the transition rates in (12.50) simplify to

$$\hat{\lambda}_{01} \sim N u_1 \psi(\rho u_2^{1/2}), \quad \hat{\lambda}_{02} = N R(\rho) u_1 u_2^{1/2}, \quad \hat{\lambda}_{12} = N u_2, \tag{12.66}$$

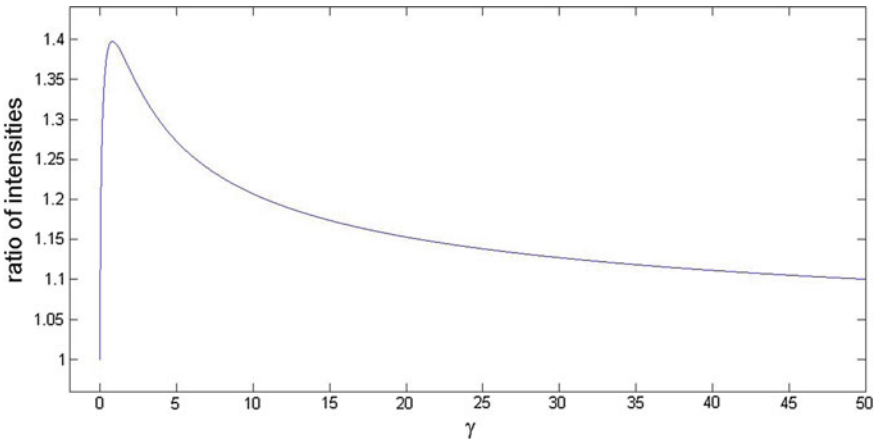

Fig. 12.1 Plot of the ratio $\xi(\gamma)$ between the approximate and exact asymptotic rates of the exponential limit distribution for the waiting time T_2 until the second mutation gets fixed in model with no backward mutations and neutral alleles ($s_1 = s_2 = 1$, i.e. $\rho = 0$ in (12.51)). The argument γ is the normalized rate (12.57) at which the second mutation occurs. It can be shown that $\xi(\gamma) > 1$ for all $\gamma > 0$, with $\xi(\gamma) \to 1$ as either $\gamma \to 0$ or $\gamma \to \infty$, and the maximum value $\xi(\gamma) = 1.40$ is attained for $\gamma = 0.82$

where

$$\psi(\rho u_2^{1/2}) = \begin{cases} 0, & \rho < 0, \\ 1/N, & \rho = 0, \\ \rho u_2^{1/2}, & \rho > 0, \end{cases}$$

relies on the asymptotic approximation of $\beta(s_1) = \beta(1 + \rho u_2^{1/2})$ defined in (12.33), when s_1, the selective fitness of allele 1, is given by (12.51). If follows from (12.4) that $\max(\hat{\lambda}_{01}, \hat{\lambda}_{02}) \ll \hat{\lambda}_{12}$ as $N \to \infty$, so that $I_{\text{long}} = \{0\}$. Regarding $\hat{\lambda}_{01}$ and $\hat{\lambda}_{02}$, their asymptotic ordering will depend on s_1. We have that $\hat{\lambda}_{01} \ll \hat{\lambda}_{02}$ if $\rho \leq 0$, whereas $\hat{\lambda}_{01}$ and $\hat{\lambda}_{02}$ are of the same order when $\rho > 0$. This means that 1 is an asymptotic state with a short waiting time when $\rho > 0$ ($I_{\text{short}} = \{1\}$), whereas it is a non-asymptotic state when $\rho \leq 0$ ($I_{\text{nas}} = \{1\}$). We follow the remark below Theorem 12.1 in (12.29), and let $\hat{\mu}_{\min} = N u_1 u_2^{1/2}$ be the asymptotic rate until allele m gets fixed in the population, which differs from μ_0 by a conveniently chosen constant. The normalized rates of leaving states 0 and 1, on the new time scale determined by $\hat{\mu}_{\min}$, are $\kappa_0 = 1(\rho > 0)\rho + R(\rho)$, $\kappa_1 = \infty$, where $1(A)$ is the indicator function for the event A (that is, it equals 1 if A occurs and 0 it if does not). The asymptotic probability of tunneling from 0 to 2, is $\pi_{02} = \frac{R(\rho)}{1(\rho>0)\rho+R(\rho)}$, and the other nonzero asymptotic transition probabilities from non-absorbing states, are $\pi_{01} = 1 - \pi_{02}$ and $\pi_{12} = 1$. This gives a rescaled intensity matrix Σ that equals (12.59). Therefore, when the mutation rate of allele 2 is large, as in (12.65), it will always become fixed in the population through tunneling $0 \to 2$ when allele 1 is selectively neutral or deleterious compared to allele 0 ($\rho \leq 0$). On the other hand, when allele 1 has higher fitness

than allele 0 ($\rho > 0$), it is possible to reach allele 2 either by tunneling, or by first having allele 1 fixed. In the latter case, the subsequent waiting time for allele 2 to spread is negligible. Formula (12.49) suggests a limit distribution

$$[1(\rho > 0)\rho + R(\rho)]\, N u_1 u_2^{1/2} \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.67}$$

as $N \to \infty$ for the total waiting time T_2 until allele 2 takes over the population. However, we only expect (12.67) to be correct when $\rho \leq 0$, since (12.44) is violated when $\rho > 0$, due to the competition between alleles 1 and 2 to take over the population at first. When $\rho \leq 0$, formula (12.67) agrees with Theorem 4 in Durrett and Schmidt [19]. In particular, when $\rho = 0$ we find that

$$N u_1 u_2^{1/2} \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.68}$$

as $N \to \infty$, since $R(0) = 1$. This agrees with a result given on pp. 231–232 of Nowak [52], and (12.68) is also a special case of Theorem 1 of Durrett et al. [20].

12.7.1.2 Backward Mutations, and Final Allele Is Neutral

In this subsection we will make three assumptions that increase the difficulty of having allele 2 fixed in the population, so that the waiting time T_2 gets longer compared to Sect. 12.7.1.1. First, we allow for backward mutations ($v > 0$), second, we assume that the fitness of the final allele 2 is the same as for allele 0 ($s_2 = 1$), and third, the intermediate allele 1 does not have a selective advantage in comparison to the other two alleles, so that $\rho \leq 0$ in (12.51). In order to avoid too many parameters of the model, we will also assume that the two forward mutation rates are identical, i.e., $u_1 = u_2 = u$. We will not consider the case when the forward mutation rate is small in comparison to the backward rate ($u = o(v)$), since (12.18) is violated then. Formally, the expected value of the limit distribution (12.27) is infinite when $u = o(v)$, since the asymptotic intensity matrix Σ_0 of fixation rates is not invertible. This is due to the fact that when backward mutations are frequent, they will effectively block the opportunities for allele 2 to spread to the whole population. In order to handle such a scenario we need to generalize Theorem 12.1 and let $\mu_{\min}^{-1}$ be determined by the asymptotic growth rate in (12.10). Here, we will therefore confine ourselves to scenarios where the backward mutation rate satisfies $v = Cu$ for some constant $C > 0$.

The above mentioned assumptions imply that the intensities (12.50) at which new alleles get fixed, simplify to

$$\begin{gathered}\hat{\lambda}_{01} = N u \beta (1 + \rho u^{1/2}),\ \ \hat{\lambda}_{02} = R(\rho) u^{3/2},\ \ \hat{\lambda}_{10} = NCu\beta(1 - \rho u^{1/2}), \\ \hat{\lambda}_{12} = N u \beta (1 - \rho u^{1/2}).\end{gathered} \tag{12.69}$$

It turns out that the asymptotic properties of T_2 depend on the size of the mutation rates, and we will look at three different scenarios.

Case 1. Small mutation rate. If

$$u = o(N^{-2}), \tag{12.70}$$

as $N \to \infty$, then (12.69) simplifies to

$$\hat{\lambda}_{01} \sim u, \quad \hat{\lambda}_{02} \sim R(\rho)u^{3/2}, \quad \hat{\lambda}_{10} \sim Cu, \quad \hat{\lambda}_{12} \sim u, \tag{12.71}$$

so that the tunneling rate $\hat{\lambda}_{02} \ll \min(\hat{\lambda}_{01}, \hat{\lambda}_{10}, \hat{\lambda}_{12})$ can be ignored, whereas the other three fixation rates $\hat{\lambda}_{01}, \hat{\lambda}_{10}$, and $\hat{\lambda}_{12}$ are of the same order. This implies that the two non-absorbing states are asymptotic, and they both contribute to the total waiting time ($I_{\text{long}} = \{0, 1\}$), with $\hat{\mu}_0 \sim \hat{\lambda}_{01} = u$ and $\hat{\mu}_1 = \hat{\lambda}_{10} + \hat{\lambda}_{12} = (C + 1)u$. Putting $\hat{\mu}_{\min} = u$, we find that the normalized rates of leaving states 0 and 1 are $\kappa_0 = 1$ and $\kappa_1 = C + 1$ respectively, whereas the nonzero asymptotic transition probabilities from the non-absorbing states are $\pi_{01} = 1, \pi_{10} = C/(C + 1)$, and $\pi_{12} = 1/(C + 1)$. This gives a matrix

$$\Sigma = \begin{pmatrix} -1 & 1 & 0 \\ C & -(C+1) & 1 \\ 0 & 0 & 0 \end{pmatrix} \tag{12.72}$$

of rescaled fixation rates. Formula (12.49) implies a limit distribution

$$u \times T_2 \xrightarrow{\mathcal{L}} \text{PD}((1, 0), \Sigma_0) \tag{12.73}$$

of the waiting time for allele 2 to take over the population as $N \to \infty$. In particular, without backward mutations ($C = 0$), we find that T_2 has an asymptotic gamma distribution

$$u \times T_2 \xrightarrow{\mathcal{L}} \Gamma(2, 1), \tag{12.74}$$

where 2 is the form parameter and 1 the intensity parameter. Since the form parameter is integer valued, the limit is also referred to as an Erlang distribution. Notice that (12.74) is a special case of Corollary 12.1, with $\kappa_0 = \kappa_1 = 1$.

Case 2. Intermediate sized mutation rate. Suppose $u = \frac{\gamma}{N^2}$ for some positive constant γ. The fixation intensities in (12.50) then simplify to

$$\begin{aligned} &\hat{\lambda}_{01} \sim u\rho\gamma^{1/2}/(1 - \exp(-\rho\gamma^{1/2})), \quad \hat{\lambda}_{02} = R(\rho)u^{3/2}, \\ &\hat{\lambda}_{10} \sim Cu\rho\gamma^{1/2}/(\exp(\rho\gamma^{1/2}) - 1), \quad \hat{\lambda}_{12} \sim u\rho\gamma^{1/2}/(\exp(\rho\gamma^{1/2}) - 1), \end{aligned} \tag{12.75}$$

as $N \to \infty$. The distribution of the waiting time T_2 turns out to be similar to Case 1, with the difference that the selection parameter ρ will have an asymptotic impact. As in Case 1, the tunneling from allele 0 to 2 can be ignored

($\hat{\lambda}_{02} \ll \min(\hat{\lambda}_{01}, \hat{\lambda}_{10}, \hat{\lambda}_{12})$), whereas the other three fixation rates $\hat{\lambda}_{01}$, $\hat{\lambda}_{10}$, and $\hat{\lambda}_{12}$ have the same order of magnitude. Therefore, both non-absorbing states are asymptotic, with a long waiting time ($I_{\text{long}} = \{0, 1\}$). Following the remark below Theorem 12.1, we standardize the time scale with an appropriately chosen constant, so that $\hat{\mu}_{\min} = u$ has a simple form. On the new time scale, the intensities to leave states 0 and 1 are $\kappa_0 = \rho\gamma^{1/2}/(1 - \exp(-\rho\gamma^{1/2}))$, $\kappa_1 = (C+1)\rho\gamma^{1/2}/(\exp(\rho\gamma^{1/2}) - 1)$ respectively. Since the nonzero transition probabilities π_{ij} of jumping between various fixation states, are the same as in Case 1, the rates of fixation between all pairs of states, after the time transformation, are

$$\Sigma = \begin{pmatrix} -\kappa_0 & \kappa_0 & 0 \\ \frac{C}{C+1}\kappa_1 & -\kappa_1 & \frac{1}{C+1}\kappa_1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{12.76}$$

It follows from formula (12.49) that the asymptotic distribution for the total waiting time to reach allele 2, is given by

$$u \times T_2 \xrightarrow{\mathcal{L}} \text{PD}\left((1,0), \Sigma_0\right) \tag{12.77}$$

as $N \to \infty$. In particular, when there are no backward mutations ($C = 0$), (12.77) simplifies to

$$u \times T_2 \xrightarrow{\mathcal{L}} \kappa_0^{-1} X_0 + \kappa_1^{-1} X_1. \tag{12.78}$$

This is a special case of Corollary 12.1, with X_0 and X_1 two independent and exponentially distributed random variables with expected value 1. Notice also that Case 1 is essentially a $\rho \to 0$ limit of Case 2.

The expected value of the limit distribution of $u \times T_2$, on the right hand side of (12.77), has an explicit form. Using formula (12.10) for the expected value of a phase-type distribution, and putting $x = \rho\gamma^{1/2}$, we find that

$$\begin{aligned} u \times E(T_2) &\sim -(1,0)\Sigma_0^{-1}(1,1)^T \\ &= \begin{cases} 2 + C, & \rho = 0, \\ \left[(e^x - 1) + (1 - e^{-x})(1 + C)\right]/x, & \rho < 0, \end{cases} \end{aligned} \tag{12.79}$$

increases linearly with the backward rate C. In Fig. 12.2 we have plotted $u \times E(T_2)$ as a function of C for various values of the selection parameter ρ, and validated the accuracy of (12.79) with simulations.

Further details for the neutral case ($x = 0$) are given in Fig. 12.3, where the density function f_{T_2} of T_2 based on (12.9) is compared with simulation based histograms, for different values of C. Whereas f_{T_2} is gamma distributed for $C = 0$, it can be seen that its form approaches an exponential density as C grows.

Case 3. Large mutation rate. Assume that the mutation rate and the selection coefficient of allele 1 satisfy

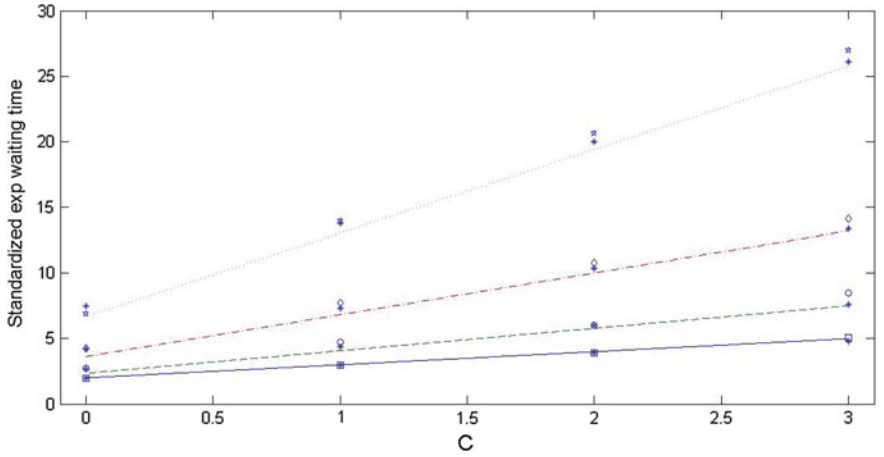

Fig. 12.2 Plot of the rescaled expected waiting time $u \times E(T_2)$, for a model with $m = 2$, forward mutation rates $u_1 = u_2 = u = \gamma/N^2$, and backward mutation rate $v_0 = Cu$. The lines are based on the approximate formula (12.79), and shown as functions of C. All lines have $s_2 = 1$, but the value of $s_1 = 1 + \rho\sqrt{u}$ varies. The intermediate allele is either neutral $\rho = 0$ (solid line), or has a selective disadvantage with $\rho = -1/\gamma^{1/2}$ (dashed line), $\rho = -2/\gamma^{1/2}$ (dash-dotted line), and $\rho = -3/\gamma^{1/2}$ (dotted line). Result from 1000 simulations, for a population of size $N = 100$ with $\gamma = 1$, are shown for $\rho = 0$ (squares), $\rho = -1/\gamma^{1/2}$ (circles), $\rho = -2/\gamma^{1/2}$ (diamonds), and $\rho = -3/\gamma^{1/2}$ (pentagrams). The parameters of the simulation algorithm are $N_c = 10$ and $\varepsilon = 0.2$ (see Appendix A). The simulation based estimates are also compared with the more accurate analytical solution (stars) based on (12.10) and (12.35)

$$\begin{aligned} &u \gg N^{-2}, \\ &\rho < 0, \end{aligned} \tag{12.80}$$

respectively. (If $\rho = 0$, it turns out that the asymptotics of T_2 is identical to Case 1.) The fixation rates in (12.50) then simplify to

$$\begin{aligned} \hat{\lambda}_{01} &= 0, \\ \hat{\lambda}_{02} &= R(\rho)u^{3/2}, \\ \hat{\lambda}_{10} &\sim -NC\rho u^{3/2}, \\ \hat{\lambda}_{12} &\sim -N\rho u^{3/2} \end{aligned} \tag{12.81}$$

as $N \to \infty$. We notice that $\hat{\lambda}_{01} \ll \hat{\lambda}_{02} \ll \min(\hat{\lambda}_{10}, \hat{\lambda}_{12})$. This implies that there is one asymptotic state $I_{\text{long}} = \{0\}$ with a long waiting time, and one non-asymptotic state $I_{\text{nas}} = \{1\}$. Time is therefore rescaled according to $\hat{\mu}_{\min} = \hat{\mu}_0 \sim \hat{\lambda}_{02} = R(\rho)u^{3/2}$, so that $\kappa_0 = 1$ and $\kappa_1 = \infty$ are the rescaled rates of leaving states 0 and 1, whereas $\pi_{02} = 1$, $\pi_{10} = C/(C+1)$, and $\pi_{12} = 1/(C+1)$ are values of the three nonzero transition probabilities from non-absorbing states. The matrix of standardized fixation rates is

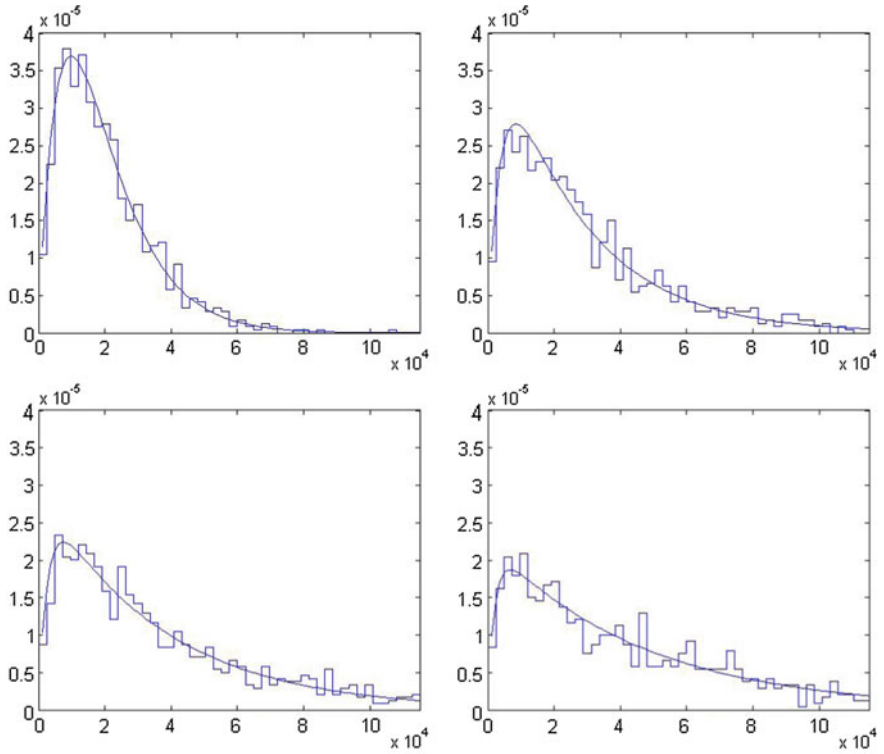

Fig. 12.3 Density functions (12.9) of the waiting time T_2 for a model with $N = 100$ individuals and $m = 2$ selectively neutral coordinated mutations ($s_1 = s_2 = 1$). The forward mutation rates are $u_1 = u_2 = 1/N^2$, whereas the backward mutation rates are $v_0 = C/N^2$ and $v_1 = 0$. The four graphs have $C = 0$ (upper left), $C = 1$ (upper right), $C = 2$ (lower left), and $C = 3$ (lower right), corresponding to the four simulations of Fig. 12.2 that are marked with squares. Shown in each plot is also a histogram from 1000 simulations, with parameters $N_c = 10$ and $\varepsilon = 0.2$ (see Appendix A). The estimated coefficients of variation $\sqrt{\text{Var}(T_2)}/E(T_2)$ from these four simulations are 0.704, 0.882, 0.941, and 0.965. This agrees well with the coefficients of variation of the density functions, which are $1/\sqrt{2} = 0.707$ for $C = 0$, and 1 in the limit as $C \to \infty$

$$\Sigma = \begin{pmatrix} -1 & 0 & 1 \\ C \times \infty & -(C+1) \times \infty & \infty \\ 0 & 0 & 0 \end{pmatrix}, \tag{12.82}$$

where ∞ of the second row should be interpreted as a limit. However, since the selective disadvantage of allele 1 is so large compared to alleles 0 and 2, allele 1 will never be fixed in a large population, and the second row of Σ will have no impact. Therefore, the only way of reaching allele 2 is through stochastic tunneling from allele 0. Formula (12.49) gives the limit distribution

$$R(\rho)u^{3/2} \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.83}$$

as $N \to \infty$ for the waiting time of allele 2 to get fixed. This is also a special case of Corollary 12.1.

12.7.2 Arbitrary Number of Coordinated Mutations

In this subsection we look at models with an arbitrary number m of coordinated mutations and number $m+1$ of alleles. We will consider two different kinds of models. The first one has no backward mutations, but the forward mutations have to arrive in a pre-specified order. The second model incorporates backward mutations, but the forward mutations may enter the population in any order.

12.7.2.1 Equal Forward Mutation Rates, No Backward Mutations

Assume there are no backward mutations ($v_0 = \cdots = v_{m-1} = 0$), and that forward mutations have to appear in a pre-determined order with identical mutation rates, i.e.

$$u_1 = \cdots = u_m = u. \tag{12.84}$$

We will also assume that all intermediate alleles are neutral or deleterious with the same selective fitness

$$s_1 = \cdots = s_{m-1} = s \le 1, \tag{12.85}$$

where

$$s = 1 + \rho u^{1-2^{-(m-1)}} \tag{12.86}$$

for some fixed constant $\rho \le 0$, not depending on N, and that the final allele m has a high fitness ($s_m = \infty$).

With these assumptions, formulas (12.35)–(12.36) for the fixation rates between different pairs of alleles simplify to $\hat{\lambda}_{ij} = 0$ when $j < i$, and to

$$\hat{\lambda}_{ij} \sim \begin{cases} NR(\rho)^{I(j=m)} u^{2-2^{-(j-1)}} \beta (1 + \rho u^{1-2^{-(m-1)}})^{I(j<m)}, & i = 0, \\ N u^{2-2^{-(j-i-1)}} (1/N)^{I(j<m)}, & i = 1, \ldots, m-1 \end{cases} \tag{12.87}$$

when $j > i$, where $I(\cdot)$ is the indicator function. In (12.87) we simplified the expressions for the terms $R(\rho_{ilj})$ on the right hand side of Eq. (12.36), for all i, l, j with $0 \le i < l < j \le m$. More specifically, we utilized that

$$\begin{aligned} \rho_{01m} &\sim \rho, \\ \rho_{0lj} &= o(1), (l, j) \neq (1, m), \\ \rho_{ilj} &= 0, \quad i = 1, \ldots, m-2, \end{aligned} \tag{12.88}$$

which, in view of (12.41), implies $R(\rho_{01m}) \sim R(\rho)$ and $R(\rho_{ilj}) \sim 1$ for the other terms of (12.36). In order to motivate (12.88), we use formulas (12.37) and (12.85)–(12.86) to find that

$$\begin{aligned} &1 + \rho_{ilj} u^{1/2} r_{ilj}^{1/2} = \tfrac{s_l}{s_i} \\ &= \begin{cases} 1 + \rho u^{1-2^{-(m-1)}}, & i = 0,\ l = 1, \ldots, j-1, \\ 1, & i = 1, \ldots, m-2,\ l = i+1, \ldots, j-1. \end{cases} \end{aligned} \tag{12.89}$$

When $i > 0$, (12.88) follows immediately from (12.89). When $i = 0$, we find that

$$\rho_{ilj} = \rho u^{1/2 - 2^{-(m-1)}} r_{ilj}^{-1/2}, \tag{12.90}$$

and therefore we also need to find expressions for r_{ilj}. To this end, we use formula (12.143) of Appendix C to deduce

$$\begin{aligned} r_{0lj} &= \Theta(u^{1-2^{-(j-l-1)}}), \\ r_{01m} &\sim u^{1-2^{-(m-2)}}. \end{aligned} \tag{12.91}$$

Then we insert (12.91) into (12.90) and use formula (12.4) to notice that $u \to 0$ as $N \to \infty$, in order to prove the upper two equations of (12.88).

Having established formula (12.87) for the transition rates between different fixed states, we will next investigate which jumps from state $i \leq m-2$ that are possible when N gets large. It follows from (12.87) that the transition rates from i to the intermediate states $i+1, \ldots, m-1$ are related as

$$\hat{\lambda}_{i,i+1} \gg \hat{\lambda}_{i,i+2} \gg \cdots \gg \hat{\lambda}_{i,m-1} \tag{12.92}$$

when $N \to \infty$, and therefore it is not possible to have a direct transition from state i to any of $j = i+2, \ldots, m-1$. This can also be deduced directly from formula (12.45). A transition from i to $i+2 \leq j < m$ is not possible asymptotically, since s_j is not larger than $\max(s_{i+1}, \ldots, s_{j-1})$.

Consequently, it is only possible for a population with allele i, to transfer either to a population with $i+1$ alleles, or to one in which the final allele m has been fixed. Therefore, the rate of leaving state i is of the order

$$\begin{aligned} \hat{\mu}_i &= \textstyle\sum_{j=i+1}^{m} \hat{\lambda}_{ij} \\ &= \Theta\left(\max(\hat{\lambda}_{i,i+1}, \hat{\lambda}_{im})\right) \end{aligned} \tag{12.93}$$

Table 12.1 Some possible scenarios for m coordinated mutations when all mutation rates (12.84) are identical, and the selective fitness (12.85) is the same for all alleles $i < m$. The dots indicate successive transitions $i \to i+1$ between neighboring alleles

Case	Scenario	Transitions	Mutation rate u
3	0	$0 \to m$	Large
2	$1 \le n \le m-2$	$0 \to \cdots \to n \to m$	Intermediate
1	$m-1$	$0 \to \cdots \to m$	Small

as $N \to \infty$. The asymptotic properties of the waiting time T_m until allele m gets fixed, will depend on which of the rates on the right hand side of (12.93) that dominate as $N \to \infty$ among the asymptotic states. We will consider m different scenarios, numbered as $n = 0, \ldots, m-1$, where Scenario n is characterized by a set

$$I_{\text{as}} = \{0, \ldots, n\} \tag{12.94}$$

of asymptotic states. These scenarios can be divided into three groups, depending on the size of the mutation rate u (see Table 12.1). As a general rule, the larger the mutation rate is, the earlier stochastic tunneling will kick in and drive the population towards its final state, where allele m has been fixed.

Case 1. Small mutation rate. Suppose

$$u = o(N^{-2}) \tag{12.95}$$

as $N \to \infty$, so that the rates of fixation between different pairs of alleles in (12.87) simplify to

$$\hat{\lambda}_{ij} \sim NR(\rho)^{I(i=0,j=m)} u^{2-2^{-(j-i-1)}} (1/N)^{I(j<m)}, \tag{12.96}$$

for $i = 0, \ldots, m-1$ and $i < j \le m$. We used (12.33) to simplify the fixation probability in (12.87) to

$$\beta(1 + \rho u^{1-2^{-(m-1)}}) \sim N^{-1}, \tag{12.97}$$

since (12.95) implies $u^{1-2^{-(m-1)}} = o(N^{-1})$.

Recall from (12.93) that asymptotically, we only have to consider transitions from i to $i+1$ and m. We deduce from formula (12.96) that the rates of leaving the non-absorbing states are $\hat{\mu}_i \sim \hat{\lambda}_{i,i+1} = u$ for $i = 0, 1, \ldots, m-2$, and $\hat{\mu}_{m-1} = \hat{\lambda}_{m-1,m} = Nu$. This corresponds to Scenario $m-1$ in (12.94), but only the first $m-1$ asymptotic states will contribute to the overall waiting time, so that $I_{\text{long}} = \{0, \ldots, m-2\}$ and $I_{\text{short}} = \{m-1\}$. Rescaling time by a factor $\hat{\mu}_{\min} = u$, we find that the normalized rates of leaving state i are $\kappa_i = 1$ for $i \in I_{\text{long}}$ and $\kappa_{m-1} = \infty$. Since the nonzero asymptotic transition probabilities for jumps from non-absorbing allelic states are $\pi_{i,i+1} = 1$ for $i = 0, \ldots, m-1$, the conditions of Corollary 12.1 are satisfied. It follows from formula (12.31) that

$$u \times T_m \xrightarrow{\mathcal{L}} \Gamma(m-1,1) \tag{12.98}$$

as $N \to \infty$, in agreement with Case 1 in Theorem 2 of Schweinsberg [60].

Case 2. Intermediate mutation rate. Let $n \in \{1, \ldots, m-2\}$ be a fixed number, and assume that the mutation rate has size

$$N^{-1/(1-2^{-(m-n-1)})} \ll u \ll N^{-1/(1-2^{-(m-n)})} \tag{12.99}$$

as $N \to \infty$. The transition rates $\hat{\lambda}_{ij}$ and the fixation probability satisfy (12.96) and (12.97), also for Case 2. It follows that the rate at which allele i is lost in a large population is $\hat{\mu}_i \sim \hat{\lambda}_{i,i+1} = u$ for $i = 0, \ldots, n-1$, and $\hat{\mu}_n \sim \hat{\lambda}_{nm} = Nu^{2-2^{-(m-n-1)}} \gg u$ for $i = n$. The nonzero asymptotic transition probabilities from non-absorbing states are therefore $\pi_{i,i+1} = 1$ for $i = 0, \ldots, n-1$, and $\pi_{nm} = 1$. This corresponds to Scenario n in (12.94), where the first n asymptotic states contribute to the overall waiting time ($I_{\text{long}} = \{0, \ldots, n-1\}$), the remaining asymptotic state n does not ($I_{\text{short}} = \{n\}$), and the other non-absorbing states are non-asymptotic ($I_{\text{nas}} = \{n+1, \ldots, m-1\}$). If time is standardized by $\hat{\mu}_{\min} = u$, the rescaled rates of leaving state i are $\kappa_i = 1$ for $i \in I_{\text{long}}$, and $\kappa_n = \infty$. Since the conditions of Corollary 12.1 are satisfied, we apply formula (12.31) and deduce an asymptotic distribution

$$u \times T_m \xrightarrow{\mathcal{L}} \Gamma(n,1) \tag{12.100}$$

as $N \to \infty$ for the waiting time until allele m gets fixed in the population. This corresponds to Case 2 of Theorem 2 in Schweinsberg [60].

Case 3. Large mutation rate. Assume that

$$N^{-1/(1-2^{-(m-1)})} \ll u \ll N^{-1/(1-2^{-m})} \tag{12.101}$$

as $N \to \infty$. It can be seen that the fixation probability satisfies (12.97) when $\rho = 0$, whereas $\beta(1+\rho u^{1-2^{-(m-1)}}) = o(N^{-1})$ when $\rho < 0$. In any case, it follows from (12.87) that $\hat{\lambda}_{01} \le u \ll \hat{\lambda}_{0m} = NR(\rho)u^{2-2^{-(m-1)}}$. Transitions from 0 will therefore be to state m when N is large, so that the rate of leaving state 0 is of the order $\hat{\mu}_0 \sim \hat{\lambda}_{0m}$. This corresponds to Scenario 0 in (12.94), with one single state $I_{\text{long}} = \{0\}$ that contributes to the waiting time T_m asymptotically, and since $\pi_{0m} = 1$, all other non-absorbing states are non-asymptotic ($I_{\text{nas}} = \{1, \ldots, m-1\}$). With $\hat{\mu}_{\min} = \hat{\mu}_0$, the normalized rate of leaving state 0 is $\kappa_0 = 1$. Since Corollary 12.1 is satisfied, we deduce from (12.31) that the waiting time for the m:th mutant to get fixed, has a limiting distribution

$$NR(\rho)u^{2-2^{-(m-1)}} \times T_m \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.102}$$

as $N \to \infty$. This results generalizes Case 3 of Theorem 2 in Schweinsberg [60] from the neutral case $\rho = 0$ to $\rho \le 0$.

Boundary scenarios. As in Schweinsberg [60], it is possible to consider $m-1$ additional asymptotic scenarios for the mutation rate, which can be interpreted as boundaries between the m scenarios of Table 12.1. For simplicity, we confine ourselves to the neutral case $\rho = 0$. Suppose $n \in \{0, \ldots, m-2\}$. As a boundary between scenarios n and $n+1$ of Table 12.1, we assume that the mutation rate satisfies

$$u = \frac{\gamma}{N^{-1/(1-2^{-(m-n-1)})}} \tag{12.103}$$

as $N \to \infty$, for some constant $\gamma > 0$. In this case the population dynamics starts with n fixation events $0 \to 1 \to \cdots \to n$. Then in the next step there is competition between fixation $n \to n+1$ and tunneling $n \to m$. If $n+1$ gets fixed, in the next step there will be a much faster transition $n+1 \to m$ that does not contribute to the overall waiting time. Therefore, among the asymptotic states $I_{\text{as}} = \{0, \ldots, n+1\}$, only those in $I_{\text{long}} = \{0, \ldots, n\}$ contribute asymptotically to T_m.

In more detail, combining the arguments for Cases 1–3 above, it can be seen that the rates of leaving state i is $\hat{\mu}_i \sim \hat{\lambda}_{i,i+1} = u$ for $i = 0, \ldots, n-1$, whereas $\hat{\mu}_n \sim \hat{\lambda}_{n,n+1} + \hat{\lambda}_{nm} = u + Nu^{2-2^{-(m-n-1)}} = (1 + \gamma^{1-2^{-(m-n-1)}})u$ for state n. We then transform the time scale by $\hat{\mu}_{\min} = u$, and find that the normalized rates are $\kappa_i = 1$ of leaving states $i = 0, \ldots, n-1$, $\kappa_n = 1 + \gamma^{1-2^{-(m-n-1)}}$ to leave state n, and it is $\kappa_{n+1} = \infty$ to leave state $n+1$. Therefore, Theorem 12.2 suggests a limit distribution

$$u \times T_m \xrightarrow{\mathcal{L}} X_0 + \cdots X_{n-1} + \frac{1}{1 + \gamma^{1-2^{-(m-n-1)}}} X_n \tag{12.104}$$

as $N \to \infty$, for the waiting time until allele m gets fixed, where $X_0, \ldots, X_n$ are independent random variables with an identical distribution that is exponential with expected value 1. However, the limit distribution in (12.104) is incorrect. The reason is that regularity condition (12.44) is violated for transitions from state n. Asymptotically it is possible to either have a transition $n \to n+1$ or stochastic tunneling $n \to m$, and therefore $\pi_{n,n+1}$ and π_{nm} are both positive. The correct limit distribution for T_m is given in Theorem 3 of Schweinsberg [60]. It states that

$$u \times T_m \xrightarrow{\mathcal{L}} X_0 + \cdots X_{n-1} + \frac{1}{\chi(\gamma^{2(1-2^{-(m-n-1)})})} X_n \tag{12.105}$$

as $N \to \infty$, with $\chi(\cdot)$ defined in (12.63). We notice that (12.104)–(12.105) generalize (12.61)–(12.62), which corresponds to the special case $n = 0$ and $m = 2$. The approximate limit distribution in (12.104) has a slightly lower expected value than the correct one in (12.105), and their ratio will depend on n and $\gamma' = \gamma^{2(1-2^{-(m-n-1)})}$. In Table 12.2 we have displayed the maximal possible ratio between expected values of the correct and approximate limit distribution, as a function of n. It can be seen that this ratio quickly approaches 1 as n grows. For most values of γ' (or γ), the ratio will be even closer to 1. See also Sect. 12.8, were we introduce a method that to some extent corrects for the different expected waiting times of (12.104) and (12.105).

Table 12.2 The table refers to a model with no backward mutations and m forward mutations with equal rate ($= u$) that satisfies (12.103) for some $\gamma > 0$ and $0 \le n \le m-2$, so that a direct transition $n \to n+1$ and tunneling $n \to m$ are both possible. Displayed is the maximal possible ratio between the expected values of the correct and approximate limit distributions of the time until allele m gets fixed, in (12.105) and (12.104) respectively, as a function of the number n of transitions $0 \to \cdots \to n$ without any tunneling. The maximum ratio in the table is attained for a value of γ that depends on n. When $n = 0$, it equals the maximum of the function that is plotted in Fig. 12.1

n	Maximal ratio
0	1.398
1	1.143
2	1.088
3	1.064
4	1.050

12.7.2.2 Forward and Backward Mutations in Any Order

When forward and backward mutations are allowed to arrive in any given order, it is reasonable to identify type i with the number of mutations that have appeared in the population so far. Suppose that u and $v = Cu$ are the rates at which each single forward and backward mutation arrive. When i mutants have been fixed in the population, there are $m - i$ additional forward mutations not present in the population, and i possible types of back mutations. Consequently, $u_{i+1} = (m-i)u$, $v_{i-1} = Ciu$, for $i = 0, \ldots, m-1$. We will also assume a neutral model, so that $s_1 = \cdots = s_m = 1$. Then formulas (12.35)–(12.36) simplify to

$$\hat{\lambda}_{ij} = \begin{cases} \prod_{l=i}^{j-1} [(m-l)u]^{2^{-(l-i)}}, & j > i, \\ \prod_{l=j+1}^{i} (Clu)^{2^{-(i-l)}}, & j < i. \end{cases} \tag{12.106}$$

Since the model is neutral, the tunneling condition in (12.45) is violated for all pairs i, j of states. We may therefore disregard the possibility of tunneling, asymptotically as $N \to \infty$, so that the rate of leaving state i is of the order

$$\hat{\mu}_i \sim \hat{\lambda}_{i,i-1} + \hat{\lambda}_{i,i+1} = [m + (C-1)i]\, u,$$

for $i = 0, \ldots, m-1$. Since all $\hat{\mu}_i$ have the same asymptotic order, it follows that all non-absorbing states are asymptotic with a long waiting time ($I_{\text{long}} = \{0, \ldots, m-1\}$). It is convenient to transform the time scale by $\hat{\mu}_{\min} = u$, so that the asymptotic rescaled intensity matrix in (12.28) has elements

$$\Sigma_{ij} = \begin{cases} m-i, & j = i+1, \\ Ci, & j = i-1, \\ 0, & |j-i| \ge 2, \\ -[m + (C-1)i], & j = i, \end{cases} \tag{12.107}$$

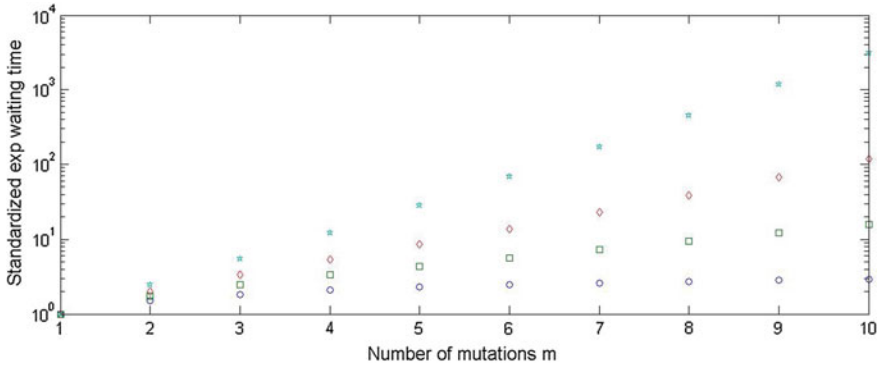

Fig. 12.4 Plot of standardized asymptotic expected waiting time $u \times E(T_m)$, according to formula (12.108), as a function of the number of required mutations m. The forward and backward mutations may appear in any order. The forward mutation rate per allele is u, and the symbols correspond to different rates $v = Cu$ of backward mutations per allele, with circles ($C = 0$), squares ($C = 0.5$), diamonds ($C = 1$), and pentagrams ($C = 2$)

for the rows that correspond to non-absorbing states ($i < m$). Combining (12.10) and (12.49), we find that the expected waiting time is given by

$$E(T_m) \sim \tilde{e}_0 \Sigma_0^{-1} 1 \times u^{-1}, \tag{12.108}$$

asymptotically as $N \to \infty$. In Fig. 12.4 we plotted the expected waiting time in (12.108) as a function of m, for different values of C. While $E(T_m)$ increases quite slowly with m in absence of backward mutations, there is a dramatic increase of $E(T_m)$ for positive C as the number m of required mutations increases. In Appendix C we derive an explicit formula for the asymptotic approximation (12.108) of $E(T_m)$. It follows from this derivation that (12.108) can be approximated by the simpler but somewhat less accurate expression

$$u \times E(T_m) \sim \begin{cases} \log(m) + 0.577, & C = 0, \\ (1 + C)^m/(Cm), & C > 0, \end{cases} \tag{12.109}$$

when C is fixed and m gets large. Formula (12.109) underscores the staircase behavior of the expected waiting time with increasing m when $C > 0$. This behavior would be even more dramatic if the intermediate states had a selective disadvantage ($s_i < 1$ for $i = 1, \ldots, m - 1$), cf. Figure 2 of Axe [2].

12.8 Some Improvements of the Asymptotic Waiting Time Theory

The practical implication of Theorem 12.2 is to approximate the distribution of the waiting time T_m until the mth mutant gets fixed. We expect this distribution to be accurate for large populations with mutation rates (12.4)–(12.5) smaller than the

inverse population size. Second, according to (12.44) there should not be competition between different alleles to get fixed. That is, for any allele i there should be at most one forward rate of fixation from i, and at most one backward rate of fixation from i, that dominate. Third, the time it takes for alleles to get fixed should be asymptotically negligible because of (12.47). In this section we will highlight some possible improvement of formula (12.10) for the expected waiting time $E(T_m)$ based on transition rates (12.35), when some of these conditions fail. Our discussion is not at all complete, but we hope it will open up for further research.

We will first revisit Case 4 of Sect. 12.7.1.1, that is, a model with $m = 2$ mutants and a large second forward mutation rate u_2. We will see what happens when the first forward mutation rate u_1 is no longer of smaller order than the inverse population size. The following result, which generalizes Theorem 1 of Durrett et al. [20], is proved in Appendix C:

Theorem 12.3 *Consider a Moran model with $m = 2$ and no backward mutations ($v_0 = 0$), where the sizes of the two forward mutation rates satisfy $Nu_1 \to a$ for some $a \geq 0$ and $N\sqrt{u_2} \to \infty$ as $N \to \infty$. Assume that the first selection coefficient $s_1 = s$ is given by (12.51) for some fixed $\rho \leq 0$, and the second one is large ($s_2 = \infty$). Let also T_2'' be the time point when the first successful mutant 2 appears in the population. Then*

$$P(NR(\rho)u_1\sqrt{u_2} \times T_2'' \geq t) \sim \exp\left(-\int_0^t h(x)dx\right) \tag{12.110}$$

as $N \to \infty$, where $R(\rho)$ is defined in (12.41), $h(x) = h(x; a, \rho)$ is a hazard function that satisfies $h(x) = 1$ when $a = 0$, and

$$h(x) = \frac{1 - \exp\left(-\frac{2\sqrt{\rho^2+4}}{\rho+\sqrt{\rho^2+4}} \times \frac{x}{a}\right)}{1 + \frac{\sqrt{\rho^2+4}+\rho}{\sqrt{\rho^2+4}-\rho} \exp\left(-\frac{2\sqrt{\rho^2+4}}{\rho+\sqrt{\rho^2+4}} \times \frac{x}{a}\right)} \tag{12.111}$$

when $a > 0$. In particular, the expected waiting time is approximated by

$$E(T_2'') \sim \left[NR(\rho)u_1\sqrt{u_2}\right]^{-1} \theta(a, \rho) = \hat{\lambda}_{02}^{-1}\theta(a, \rho), \tag{12.112}$$

where $\hat{\lambda}_{02}$ is the transition rate defined in (12.66), and

$$\theta(a, \rho) = \int_0^\infty \exp\left(-\int_0^t h(x; a, \rho)dx\right) dt.$$

In Theorems 12.1 and 12.2, we imposed conditions so that the time of tunneling and fixation were asymptotically negligible. Theorem 12.3 reveals that this is no longer the case when $u_1 = \Theta(N^{-1})$, since the waiting time T_2'' includes two parts of comparable size; the time T_2' until the first successful $0 \to 1$ mutation appears, *and* the time $T_2'' - T_2'$ of tunneling, that is, the time between the arrival of the first

successful 1 mutant and the first successful 2 mutant. It follows from the proof of Theorem 12.3 that the time T_2' until the first successful 1 mutant appears has an asymptotic exponential distribution with expected value $E(T_2') \sim \hat{\lambda}_{02}^{-1}$. Therefore, in view of (12.112), we find that tunneling *multiplies* the expected waiting time by a factor $\theta(a, \rho)$. On the other hand, we recall from Sect. 12.5 that the time it takes for allele 2 to become fixed after its first appearance, *adds* a term $\alpha(s_2) \sim E(T_2 - T_2'')$ to the expected waiting time $E(T_2) = E(T_2'') + E(T_2 - T_2'')$.

We will apply these findings as follows: Let $\hat{\lambda}_{ij}$ be the approximate fixation rates in (12.35). When $j \neq i$ we modify these rates as

$$\tilde{\lambda}_{ij} = \begin{cases} \left[\hat{\lambda}_{ij}^{-1} + \alpha(s_j/s_i)\right]^{-1}, & j \neq i, |j-i| \neq 2, \\ \left[\hat{\lambda}_{ij}^{-1}\theta(a_{i,i+1}, \rho_{i,i+1}) + \alpha(s_{i+2}/s_i)\right]^{-1}, & j = i+2, \\ \left[\hat{\lambda}_{ij}^{-1}\theta(a_{i,i-1}, \rho_{i,i-1}) + \alpha(s_{i-2}/s_i)\right]^{-1}, & j = i-2, \end{cases} \tag{12.113}$$

to take tunneling and fixation into account, where $a_{i,i+1} = Nu_{i+1}\beta(s_{i+2}/s_i)$ and $a_{i,i-1} = Nv_{i-1}\beta(s_{i-2}/s_i)$ are the size normalized rates at which new mutations appear *and* get fixed, conditionally on that tunneling is successful, whereas $s_{i+1}/s_i = 1 + \rho_{i,i+1}\sqrt{u_{i+2}}$ and $s_{i-1}/s_i = 1 + \rho_{i,i-1}\sqrt{v_{i-2}}$ are special cases of (12.37) and (12.39) for tunneling over one allele ($|j-i| = 2$). When $i = j$, we define $\tilde{\lambda}_{ii}$ so that all row sums of the matrix with elements $\tilde{\lambda}_{ij}$, are zero. The modified transition rates in (12.113) only incorporate the impact of tunneling over one allele, because it is more complicated to correct for tunneling over larger distances, and it is likely that this has less impact in many applications.

As a next step, we will correct for competition between different states to become fixed. We will confine ourselves to the case when all mutants except the last one are selectively neutral ($s_1 = \cdots = s_{m-1} = 1$), and the last mutant has a selective advantage ($s_m > 1$). It follows from this and the discussion above (12.45) that it is only possible to have competition between fixation events $i \to i+1$ and $i \to m$ for a population whose current fixed state is i. We therefore compare these two transition rates, as defined in (12.113), and denote their squared ratio by $\gamma_i = \left(\frac{\tilde{\lambda}_{im}}{\tilde{\lambda}_{i,i+1}}\right)^2$. In Appendix C we motivate that the forward transition rates in (12.113) should be modified as

$$\bar{\lambda}_{ij} = \frac{\chi\left[\gamma_i/\beta(s_m)\right]}{1+\sqrt{\gamma_i}} \times \tilde{\lambda}_{ij}, \quad j = i+1, \ldots, m, \tag{12.114}$$

where $\chi(\gamma)$ was introduced in (12.63). We put $\bar{\lambda}_{ij} = \tilde{\lambda}_{ij}$ when $j < i$, whereas the diagonal terms $\bar{\lambda}_{ii}$ are chosen so that all row sums of the matrix $\bar{\Lambda} = (\bar{\lambda}_{ij})$, are zero. When $s_m = \infty$ (so that $\beta(s_m) = 1$), we notice that the multiplicative correction factor of (12.114) is $\xi(\gamma)^{-1}$, where $\xi(\gamma)$ is the function defined in (12.64) and plotted in Fig. 12.1. Therefore, when the last mutant has high fitness, this figure tells how much the expected waiting time of a forward fixation from i will increase when competition

between fixed states $i+1$ and m is taken into account. This also agrees with formula (12.105).

Putting everything together, we define the adjusted expected waiting time as

$$E(T_m)_{\text{adj}} = \tilde{e}_0 \bar{\Lambda}_0^{-1} 1^T, \tag{12.115}$$

where $\bar{\Lambda}_0$ is a matrix containing the first m rows and the first m columns of $\bar{\Lambda}$, and adj is an acronym for adjusted. We regard (12.115) as the expected time until a semi-Markov process of allele frequencies Z_t with state space (12.6) reaches the absorbing state e_m. By this we mean that jumps between fixed states follow a Markov chain with a transition probability $-\bar{\lambda}_{ij}/\bar{\lambda}_{ii}$ from e_i to e_j. But the holding time in each state is no longer exponentially distributed, when tunneling and the time of fixation of alleles is taken into account. Although the time until a semi-Markov processes reaches an absorbing state does not have a phase-type distribution, if $-\bar{\lambda}_{ii}^{-1}$ is the expected holding time in fixed state i, formula (12.115) will still give the correct expected waiting time until the m:th mutant gets fixed.

12.8.1 One Mutation

In this subsection we consider a model with only one mutant ($m = 1$). Formulas (12.10) and (12.35) approximate the expected waiting time until fixation as

$$E(T_1) = \frac{1}{Nu_1\beta(s_1)}. \tag{12.116}$$

For a model with only two alleles, there is no tunneling and no competition between different states to become fixed. It is therefore only the expected time of a successful mutation to get fixed, that will influence the adjusted waiting time formula (12.115). It can be seen that this equation simplifies to

$$E(T_1)_{\text{adj}} = \frac{1}{Nu_1\beta(s_1)} + \alpha(s_1), \tag{12.117}$$

for a model with one single mutant. In Table 12.3, we have compared the accuracy of (12.116) and (12.117) with simulation based estimates of the expected waiting time. It can be seen that (12.117) is consistently a much more accurate approximation of the simulation based values. We also notice from this table that the smaller the mutation rate is, the smaller is the impact of the expected fixation time $\alpha(s_1)$. The general condition for asymptotic negligibility of the fixation time is (12.47). It simplifies to $\alpha(s_1) \ll [Nu_1\beta(s_1)]^{-1}$ for a model with one mutant, that is, scenarios for which the second term of (12.117) is small in comparison to the first term.

Table 12.3 Comparison between the expected waiting time formulas $E(T_1)$ and $E(T_1)_{\text{adj}}$, defined in (12.116) and (12.117) respectively, for a model with $m = 1$ mutant. The rightmost column are sample averages from 10000 simulations, with $\varepsilon = 0.04$ and $N_c = 10$ in the algorithm of Appendix A

N	u_1	s_1	$E(T_1)$	$E(T_1)_{\text{adj}}$	$\hat{E}(T_1)$
100	0.001	10/9	100.00	153.19	152.62
		2	20.00	33.95	33.36
		5	12.50	20.53	20.05
		9	11.25	18.21	17.50
		1000	10.01	15.89	15.34
1000	0.0001	10/9	100.00	198.78	197.76
		2	20.00	40.90	40.12
		5	12.50	23.99	23.52
		9	11.25	21.10	20.33
		1000	10.01	18.20	17.55
100	0.0005	10/9	200.00	253.19	250.33
		2	40.00	53.95	53.11
		5	25.00	33.03	32.72
		9	22.50	29.46	28.74
		1000	20.02	25.90	25.03
100	0.0001	10/9	1000.0	1053.2	1049.15
		2	200.0	213.95	211.63
		5	125.0	133.03	131.97
		9	112.5	119.46	117.30
		1000	100.1	105.98	105.64

12.8.2 Two Coordinated Mutations, and No Back Mutations

Here, we will revisit the model of Sect. 12.7.1.1, with two mutants ($m = 2$) and no back mutations ($v_0 = 0$). We assume that the first allele is selectively neutral ($s_1 = 1$), whereas the second one has high fitness ($s_2 = 10^5$).

In Table 12.4 we compare the accuracy of two analytical formulas for the expected waiting time until the second mutant gets fixed, with simulation based estimates. We follow the scenarios of Durrett and Schmidt [19], with different population sizes and forward mutation rates u_1 and u_2. The first expected waiting time formula is based on (12.10) and (12.35), whereas the second formula is the adjustment defined in (12.115).

It can be seen from Table 12.4 that the unadjusted expected waiting time is too low, whereas the adjusted expected waiting time is consistently much closer to the simulation based estimates. The reason for this discrepancy varies between scenarios. For those scenarios where Nu_1 is not small (Case 1–2 and Drosophila), an important feature of the adjusted formula is to incorporate the time it takes for the first successful

Table 12.4 Comparison between the expected waiting time formula $E(T_2)$ based on (12.10) and (12.35), and the adjusted waiting time formula $E(T_2)_{\text{adj}}$, based on (12.115), for a model with $m = 2$ mutants with selective fitness $s_1 = 1$ and $s_2 = 10^5$. We use the same scenarios as in Table 12.2 of Durrett and Schmidt [17], with different values of the two forward mutation rates u_1 and u_2, and no backward mutations. The quantity $\hat{E}(T_2)$ refers to a sample average from B simulations based on the algorithm of Appendix A, with the first simulation parameter ε reported in the rightmost column, and the second simulation parameter N_c set to 10

Scenario	N	Nu_1	$N\sqrt{u_2}$	$E(T_2)$	$E(T_2)_{\text{adj}}$	$\hat{E}(T_2)$	B	ε
Case 1	1000	1	10	92.6	163.2	166.9	10000	0.04
	10000			919.0	1557.8	1588	10000	0.2
Case 2	1000	1/4	10	367.7	463.8	470.5	10000	0.04
	10000			3649	4564	4644	2000	0.1
Case 3	1000	1/10	10	917.9	1051	1074	10000	0.1
	10000			9108	10434	10143	1000	0.1
Case 4	1000	1/10	4	2018	2523	2484	5000	0.1
	10000			20131	25150	26420	1000	0.2
Case 5	1000	1/10	1	5501	8053	8240	2000	0.1
	10000			55002	80471	85288	1000	0.2
Drosophila	1000	1/2	$10/\sqrt{3}$	301.0	449.2	460.5	10000	0.04
	10000			2998	4417	4440	2000	0.1

allele 1 to tunnel into allele 2. For those scenarios where $\lambda_{02}/\lambda_{01} = N\sqrt{u_2}$ is not large (Case 4–5 and Drosophila), an important fact is rather that the adjusted formula incorporates competition between alleles 1 and 2 to get fixed. On the other hand, it is not crucial, for any of the scenarios of Table 12.4 to correct for the time it takes for alleles to get fixed. This has two reasons. First, the expected fixation time of allele 2 is very short ($\alpha(10^5) \sim \log(N)$). Although the expected fixation time of allele 1 is much larger ($\alpha(1) \sim N$), for those scenarios where a transition $0 \to 1$ happens fairly often (that is, when $N\sqrt{u_2}$ is not too large, as for Case 5), the overall expected waiting time is still much larger than $\alpha(1)$.

12.9 Discussion

In this paper we analyzed the waiting time until the last of m mutations appears and gets fixed in a population of constant size without any substructure. We showed that approximately, this waiting time has a phase-type distribution whenever a fixed state model is applicable, where one genetic variant at a time dominates the population. The rationale behind this result is to approximate the dynamics of the genetic composition of the population by a continuous time Markov process with $m + 1$ states; the wildtype variant and the m mutants. We also provided a general scheme for calculating the intensity matrix of this process, and thereby obtained an explicit approximation of the waiting time distribution. Our model allows for forward and

backward mutations, with different selective fitness, to appear at different rates. Once the intensity matrix of the Markov process is known, the phase-type distribution of the waiting time automatically incorporates all the pathways towards the m:th mutant that the model allows for.

We believe the findings of this paper can be extended in several ways. First, we have provided quite a detailed sketch of proofs of main results, derived previously known results as special cases, and confirmed several others by simulations. While it is outside the scope of this paper to provide full proofs; this is an important topic for further research.

Second, we argued that our explicit approximation of the expected waiting time has the correct order of magnitude, even when some of the assumptions behind the intensity rate calculations are violated. In more detail, the transition rates between pairs of fixed states will only be correct when competition between different forward and backward transitions can be neglected asymptotically. We provided an adjustment of the expected waiting time for neutral models when such competition is present, using Theorem 3 of Schweinsberg [60] and Theorem 3 of Durrett et al. [20]. A challenging task is to generalize these results to scenarios where the mutants of the model have different selective fitness.

Third, we have assumed a homogeneous population of haploid individuals with constant size. We believe our main results can be extended to include varying population size, diploidy, and recombination, as well as geographic subdivision and other types of population structure.

Fourth, in some applications there are several possible orders in which the m mutations may arrive. This can still be handled by a fixed state population model, with a phase-type distribution for the waiting time, as in Sect. 12.7.2.2. But for some scenarios of partially ordered mutations, the state space of the Markov process has to be enlarged in order to keep track of the subset of mutations that has occurred (Gerstung and Beerenwinkel [24]).

Fifth, a challenging generalization is to derive a phase-type distribution approximation of the waiting time until m coordinated targets have been fixed in the population. For instance, each type $i \in \{1, \ldots, m\}$ could represent a sequence of DNA, which, compared to previous targets $j < i$, requires one or several additional point mutations. This would extend results in Durrett and Schmidt [18] from $m = 1$ to higher values of m.

Sixth, the results of this paper could serve as a building block in order to understand the genomewide rate of molecular evolution of m coordinated mutations. In order to obtain such a rate, a selection model has to be specified, whereby the selection coefficients of the m mutants at various loci are drawn from some multivariate distribution. This can viewed as an extension of the simulation studies in Gillespie [27] and Rupe and Sanford [55] for single mutations ($m = 1$) to larger m.

Seventh, our phase-type distribution approximation of the waiting time relies heavily on the assumption that all mutation rates are smaller than the inverse population size, in order to guarantee that successful mutations arrive so infrequently and then spread so quickly that one genetic variant at a time dominates. While this is a reasonable assumption for moderately-sized populations, it is not appropriate for

large populations where different mutations will coexist, interfere, and overlap. This includes virus, bacterial or simple eukaryotic populations, as well as large cell populations of cancer progression with diverse mutational patterns. While we adjusted for non-small mutation rates for some of these models in Sect. 12.8, it is still important to derive more general results for the waiting time of coordinated mutations in large populations. Several papers have addressed this issue, see for instance Iwasa et al. [33], Desai and Fisher [16], Beerenwinkel et al. [4], Gerstung and Beerenwinkel [24], Theorem 4 of Schweinsberg [60], and Theorem 1 of Durrett et al. [20]. It is an interesting topic of future research to generalize these results to our setting of forward and backward mutations, where the mutants have a varying selective fitness.

Finally, we have developed analytical and simulation based tools in Matlab for the waiting time of coordinated mutations, based on the results of this paper. They are freely available from the first author upon request.

Acknowledgements The authors wish to thank an anonymous reviewer for several helpful suggestions that improved the clarity and presentation of the paper.

Appendix A. A Simulation Algorithm

Recall from Sect. 12.2 that the allele frequency process Z_t of the Moran model is a continuous time and piecewise constant Markov process with exponentially distributed holding times at each state $z = (z_0, \ldots, z_m) \in \mathcal{Z}$. For all but very small population sizes, it is infeasible to simulate this process directly, since the distances between subsequent jumps are very small, of size $O_p(N^{-1})$. The τ-leaping algorithm was introduced (Gillespie [25], Li [42]) in order to speed up computations for a certain class of continuous time Markov processes. It is an approximate simulation algorithm with time increments of size τ. According to the leaping condition of Cao et al. [12], one chooses $\tau = \tau(\varepsilon)$ in such a way that

$$E\left[|Z_{t+\tau,i} - Z_{ti}||Z_{ti} = z_i\right] \leq \varepsilon z_i \tag{12.118}$$

for $i = 0, \ldots, m$ and some fixed, small number $\varepsilon > 0$, typically in a range between 0.01 and 0.1.

Zhu et al. [71] pointed out that it is not appropriate to use τ-leaping for the Moran model when very small allele frequencies are updated. For this reason they defined a hybrid algorithm that combines features of exact simulation and τ-leaping. Although most time increments are of length τ, some critical ones are shorter. Then they showed that (12.118) will be satisfied by the hybrid algorithm for a neutral model with small mutation rates, when

$$\tau \leq \varepsilon/2. \tag{12.119}$$

We will extend the method of Zhu et al. [71] to our setting, where forward and backward mutations are possible. In order to describe the simulation algorithm, we

first need to define the transition rates of the Moran model. From any state $z \in \mathcal{Z}$, there are at most $(m+1)m$ jumps $z \to z + \delta_{ij}/N$ possible, where $\delta_{ij} = e_j - e_i$, $0 \le i, j \le m$ and $i \ne j$. Each such change corresponds to an event where a type i individual dies and gets replaced by a another one of type j. Since the process remains unchanged when $i = j$, we need not include these events in the simulation algorithm. It follows from Sect. 12.2 that the transition rate from z to $z + \delta_{ij}/N$ is

$$\begin{aligned} a_{ij} &= a_{ij}(z) \\ &= z_i \times \tfrac{z_j s_j}{\sum_{k=0}^m z_k s_k}(1 - u_{j+1} - v_{j-1}) + z_i \times \tfrac{z_{j-1} s_{j-1}}{\sum_{k=0}^m z_k s_k} u_j + z_i \times \tfrac{z_{j+1} s_{j+1}}{\sum_{k=0}^m z_k s_k} v_j \\ &= \tfrac{z_i}{\sum_{k=0}^m z_k s_k}\left[z_j s_j (1 - u_{j+1} - v_{j-1}) + z_{j-1} s_{j-1} u_j + z_{j+1} s_{j+1} v_j\right], \end{aligned} \tag{12.120}$$

with $u_{m+1} = v_{-1} = z_{-1} = z_{m+1} = 0$. Let N_c be a threshold. For any given state z, define the non-critical set Ω of events as those pairs (i, j) with $i \ne j$ such that both of z_i and z_j exceed N_c/N. The remaining events (i, j) are referred to as critical, since at least one of z_i and z_j is N_c/N or smaller. The idea of the hybrid simulation method is to simulate updates of critical events exactly, whereas non-critical events are updated approximately. In more detail, the algorithm is defined as follows:

1. Set $t = 0$ and $Z_t = e_0 = z$.
2. Compute the $m(m+1)$ transition rates $a_{ij} = a_{ij}(z)$ for $0 \le i, j \le m$ and $i \ne j$.
3. Compute the set $\Omega = \Omega(z)$ of critical events for the current state z.
4. Determine the exponentially distributed waiting time $e \overset{\mathcal{L}}{\in} \mathrm{Exp}(a)$ until the next critical event occurs, where $a = \sum_{(i,j) \notin \Omega} a_{ij}$ is the rate of the exponential distribution.
5. If $e < \tau$, simulate a critical event $(I, J) \notin \Omega$ from the probability distribution $\{a_{ij}/a;\ (i, j) \notin \Omega\}$, and update the allele frequency vector as $z \leftarrow z + \delta_{IJ}/N$. Otherwise, if $e \ge \tau$, simulate no critical event and leave z intact.
6. Let $h = \min(e, \tau)$. Then simulate non-critical events over a time interval of length h, and increment the allele frequency vector as

$$z \leftarrow z + \frac{1}{N} \sum_{(i,j) \in \Omega} n_{ij} \delta_{ij},$$

   where $n_{ij} \sim \mathrm{Po}(a_{ij} h)$ are independent and Poisson distributed random variables.
7. Update date time ($t \leftarrow t + h$) and the allele frequency process ($Z_t \leftarrow z$).
8. If $z = e_m$, set $T_m = t$ and stop. Otherwise go back to step 2.

We have implemented the hybrid algorithm, with N_c and ε as input parameters and $\tau = \varepsilon/2$. When the selection coefficients s_i are highly variable, a smaller value of τ is needed though in order to guarantee that (12.118) holds.

Appendix B. The Expected Waiting Time for One Mutation

In this appendix we will motivate formula (12.34). It approximates the expected number of generations $\alpha(s)$ until a single mutant with fitness s spreads and get fixed in a population where the remaining $N-1$ individuals have fitness 1, given that such a fixation will happen and that no further mutations occur. This corresponds to a Moran model of Sect. 12.2 with $m=1$ mutant, zero mutation rates ($u_1=v_0=0$), and initial allele frequency distribution $Z_0=(1-p,p)$, where $p=1/N$. For simplicity of notation we write $Z_t=Z_{t1}$ for the frequency of the mutant allele 1.

Kimura and Ohta [38] derived a diffusion approximation of $\alpha(s)$, for a general class of models. It involves the infinitesimal mean and variance functions $M(z)$ and $V(z)$ of the allele frequency process, defined through

$$\begin{aligned} E(Z_{t+h}|Z_t=z) &= z+M(z)h+o(h),\\ \mathrm{Var}(Z_{t+h}|Z_t=z) &= V(z)h+o(h) \end{aligned}$$

as $h\to 0$. In order to apply their formula to a mutation-free Moran model, we first need to find $M(z)$ and $V(z)$. To this end, suppose $Z_t=z$. Then use formula (12.120) with $m=1$ to deduce that

$$z\to z+1/N \text{ at rate } a_{01}(z)=N(1-z)\frac{zs}{1-z+zs}, \tag{12.121}$$

whereas

$$z\to z-1/N \text{ at rate } a_{10}(z)=Nz\frac{1-z}{1-z+zs}. \tag{12.122}$$

From this it follows that

$$M(z)=\frac{1}{N}\left[a_{01}(z)-a_{10}(z)\right]=(s-1)\frac{(1-z)z}{1+z(s-1)} \tag{12.123}$$

and

$$V(z)=\frac{1}{N^2}\left[a_{01}(z)+a_{10}(z)\right]=\frac{1}{N}(1+s)\frac{(1-z)z}{1+z(s-1)}. \tag{12.124}$$

We will also need the function

$$G(z)=\exp\left(-\int_0^z\frac{2M(y)}{V(y)}dy\right)=\exp(-2Ns'z),$$

with $s'=(s-1)/(s+1)$. The formula of Kimura and Ohta [38] takes the form

$$\alpha(s)=\int_p^1\psi(z)\hat{\beta}(z)\left[1-\hat{\beta}(z)\right]dz+\frac{1-\hat{\beta}(p)}{\hat{\beta}(p)}\int_0^p\psi(z)\hat{\beta}^2(z)dz, \tag{12.125}$$

where

$$\hat{\beta}(z) = \hat{\beta}(s; z) = \frac{\int_0^z G(y)dy}{\int_0^1 G(y)dy} = \frac{1 - e^{-2Ns'z}}{1 - e^{-2Ns'}} \tag{12.126}$$

approximates the fixation probability of a mutant allele that starts at frequency $Z_0 = z$. In particular, $\hat{\beta}(1/N)$ approximates the exact probability (12.32) that one single copy of an allele with fitness s takes over a population where all other individuals have fitness 1. This diffusion approximation is increasingly accurate in the limit of weak selection ($s \to 1$).

The other function of the two integrands in (12.125), is

$$\psi(z) = \frac{2\int_0^1 G(y)dy}{V(z)G(z)} = \frac{1 - e^{-2Ns'}}{e^{-2Ns'z}} \times \frac{1 + z(s-1)}{1+s} \times \frac{1}{s'z(1-z)}. \tag{12.127}$$

In order to verify (12.34) we will approximate (12.125) separately for neutral ($s = 1$), advantageous ($s > 1$), and deleterious ($s < 1$) alleles. In the neutral case $s = 1$ we let $s' \to 0$ and find that $\hat{\beta}(z) = z$ and $\psi(z) = N/[z(1-z)]$. Inserting these functions into (12.125), we obtain an expression

$$\alpha(1) = -\frac{1}{p}\left[N(1-p)\log(1-p)\right]$$

for the expected fixation time. This is essentially the middle part of (12.34) when $p = 1/N$.

When $s > 1$, we similarly insert (12.126)–(12.127) into (12.125). After some quite long calculations, it can be shown that

$$\begin{aligned}\alpha(s) \sim{}& \frac{1+s}{s-1}\log(N) \\ &+ \frac{s}{s-1}\left[\log(2s') + \int_0^1 \frac{1-e^{-y}}{y}dy - \int_1^\infty \frac{e^{-y}}{y}dy - \frac{1}{s}\int_{2s'}^\infty \frac{e^{-y}}{y}dy\right] \\ &+ \frac{e^{-2s'}}{1-e^{-2s'}} \times \frac{1}{s-1}\int_0^{2s'} \frac{1}{y}e^y(1-e^{-y})^2 dy\end{aligned} \tag{12.128}$$

as $N \to \infty$. The first term of this expression dominates for large N, and it agrees with the lower part of (12.34).

When $s < 1$, a similar calculation yields

Table 12.5 Approximations of the expected waiting time $\alpha(s) = \alpha_N(s)$ of fixation, in units of generations, for a single mutant with selection coefficient s, in a population of size N. The columns marked Diff are based on the diffusion approximation (12.125), whereas the columns marked AsDiff are asymptotic approximations of the diffusion solution, based on the middle part of (12.34) for $s = 1$, Eq. (12.128) for $s > 1$ and Eq. (12.129) for $s < 1$. The latter two formulas only work well when $|s - 1| \gg 1/N$. They have been omitted when they depart from the diffusion solution by more than 10%

s	$N = 100$		$N = 1000$		$N = 10000$	
	Diff	AsDiff	Diff	AsDiff	Diff	AsDiff
1/5	7.38	7.39	10.84	10.85	14.30	14.30
1/2	13.62	13.67	20.57	20.58	27.48	27.48
1/1.5	20.60	20.73	32.23	32.24	43.76	43.76
1/1.1	56.42	58.92	107.06	107.27	155.61	155.63
1/1.01	98.15	–	554.51	577.34	1038.1	1040.2
1/1.001	99.48	–	985.94	–	5535.0	5710.7
1	99.50	100.00	999.50	1000.0	9999.5	10000.0
1.001	99.48	–	985.94	–	5535.0	5710.8
1.01	98.16	–	554.52	577.35	1038.1	1040.2
1.1	56.47	58.97	107.11	107.32	155.66	155.68
1.5	20.80	10.93	32.43	32.44	43.96	43.96
2	13.95	14.00	20.90	20.91	27.81	27.82
5	8.03	8.04	11.49	11.50	14.95	14.95

$$\begin{aligned}\alpha(s) \sim\ & \frac{1+s}{1-s}\log(N) \\ & + \frac{s}{1-s}\left[\log(2s'') + \int_0^1 \frac{1-e^{-y}}{y}dy - \int_1^\infty \frac{e^{-y}}{y}dy - \frac{1}{s}\int_{2s''}^\infty \frac{e^{-y}}{y}dy\right] \\ & + \frac{e^{-2s''}}{1-e^{-2s''}} \times \frac{1}{1-s}\int_0^{2s''} \frac{1}{y}e^y(1-e^{-y})^2 dy\end{aligned} \tag{12.129}$$

as $N \to \infty$, with $s'' = (1-s)/(s+1)$. The first, leading term of this formula is consistent with the upper part of (12.34). The various approximations of $\alpha(s)$ are shown in Table 12.5.

Appendix C. Sketch of Proofs of Main Results

Lemma 12.1 *Let $\{\tau_k\}_{k=0}^M$ be the fixation times of the process Z_t, defined in (12.13), and τ'_{k+1} the time points when a successful mutation first occurs between two successive fixation events ($\tau_k < \tau'_{k+1} < \tau_{k+1}$). Let also μ_i be the rate in (12.15) at which successful mutations appear in a homogeneous type i population. Then*

$$P\left(\tau'_{k+1} - \tau_k > \frac{\zeta}{\mu_i} | Z_{t_k} = e_i\right) \to \exp(-\zeta) \tag{12.130}$$

as $N \to \infty$ *for all* $\zeta > 0$ *and* $i = 0, 1, \ldots, m-1$.

Sketch of proof. Let $f_i(z) = f_{i,N}(z)$ and $b_i(z) = b_{i,N}(z)$ be the probabilities that the offspring of a type $i \in \{0, \ldots, m-1\}$ individual who mutates to $i+1$ or $i-1$ is a successful forward or backward mutation, given that the allele frequency configuration is z just before replacement occurs with the individual that dies (when $i = 0$ we put $b_0(z) = 0$). Notice in particular that $f_i = f_i(e_i)$ and $b_i = b_i(e_i)$, since these two quantities are defined as the probabilities of a successful forward or backward mutation in an environment where all individuals have type i just before the mutation, that is, when $z = e_i$.

When an individual is born in a population with allele configuration z, with probability $1 - u_{i+1} f_i(z) - v_{i-1} b_i(z)$ it is not the first successful mutation between two fixation events τ_k and τ_{k+1}, given that no other successful has occurred between these two time points. Let $0 \le t_1 < t_2 < \cdots$ be the time points when a type i individual gets an offspring, and if we choose $\{Z_t\}$ to be left-continuous, the probability of no successful mutation $i \to i \pm 1$ at time t_l, where $\tau_k < t_l < \tau_{k+1}$, is $1 - u_{i+1} f_i(Z_{t_l}) - v_{i-1} b_i(Z_{t_l})$, given that no other successful mutation has occurred so far ($\tau'_{k+1} \ge t_l$). Since the left hand side of (12.130) is the probability of no mutation $i \to i \pm 1$ being successful among those that arrive at some time point in $\mathbb{T}_i(\zeta) = \{t_l;\ \tau_k < t_l \le \tau_k + \zeta/\mu_i\}$, we find that

$$\begin{aligned} &P(\tau'_{k+1} - \tau_k > \zeta/\mu_i | Z_{\tau_k} = e_i) \\ &= E\left[\prod_{t_l \in \mathbb{T}_i(\zeta)} \left(1 - u_{i+1} f_i(Z_{t_l}) - v_{i-1} b_i(Z_{t_l})\right)\right] \\ &\approx E\left[\exp\left(-u_{i+1} \sum_{t_l \in \mathbb{T}_i(\zeta)} f_i(Z_{t_l}) - v_{i-1} \sum_{t_l \in \mathbb{T}_i(\zeta)} b_i(Z_{t_l})\right)\right], \end{aligned} \tag{12.131}$$

where expectation is with respect to variations in the allele frequency process Z_t for $t \in \mathbb{T}_i(\zeta)$.

Because of (12.4)–(12.5), with a probability tending to 1 as $N \to \infty$, Z_t will stay close to e_i most of the time in (τ_k, τ'_{k+1}), that is, all alleles $l \ne i$ will most of the time be kept at low frequencies. In order to motivate this, we notice that by definition, all mutations that arrive in (τ_k, τ'_{k+1}) are unsuccessful. It is known that the expected lifetime of an unsuccessful mutations is bounded by $C \log(N)$ for a fairly large class of Moran models with selection, where C is a constant that depends on the model parameters, but not on N (Crow and Kimura [15], Section 8.9). Since mutations arrive at rate $N(v_{i-1} + u_{i+1})$, this suggest that all alleles $l \ne i$ are expected to have low frequency before the first successful mutation arrives, if

$$C \log(N) \times N(v_{i-1} + u_{i+1}) = o(1)$$

as $N \to \infty$, i.e. if the convergence rate towards zero in (12.4)–(12.5) is faster than logarithmic. This implies that it is possible to approximate the sums on the right hand sides of (12.131) by

$$\begin{array}{l} \sum_{t_l \in \mathbb{T}_i(\zeta)} f_i(Z_{t_l}) \approx f_i |\mathbb{T}_i(\zeta)| \approx f_i N \times \zeta/\mu_i, \\ \sum_{t_l \in \mathbb{T}_i(\zeta)} b_i(Z_{t_l}) \approx b_i |\mathbb{T}_i(\zeta)| \approx b_i N \times \zeta/\mu_i, \end{array} \tag{12.132}$$

where $|\mathbb{T}_i(\zeta)|$ refers to the number of elements in $\mathbb{T}_i(\zeta)$. In the first step of (12.132), we used that $f_i(z) \to f_i$ and $b_i(z) \to b_i$ as $z \to e_i$ respectively, and therefore $f_i(Z_{t_l}) \approx f_i$ and $b_i(Z_{t_l}) \approx b_i$ for most of the terms in (12.132). In the second step of (12.132) we used that $|\mathbb{T}_i(\zeta)|$ counts the number of births of type i individuals within a time interval of length ζ/μ_i, and that each $t_{l+1} - t_l$ is approximately exponentially distributed. By the definition of the Moran model in Sect. 12.2, the intensity of this exponential distribution is approximately

$$N \times \frac{Z_{t_l i} s_i}{\sum_{j=0}^{m} Z_{t_l j} s_j} \approx N,$$

for the majority of time points t_l such that Z_{t_l} stays close to e_i. Consequently, $|\mathbb{T}_i(\zeta)|$ is approximately Poisson distributed with expected value $N\zeta/\mu_i$. We know from (12.4)–(12.5) and (12.15) that $\mu_i = o(1)$. Because this implies that $N\zeta/\mu_i \gg 1$ is large, and since the coefficient of variation of a Poisson distribution tends to zero when its expected value increases, $|\mathbb{T}_i(\zeta)|/(N\zeta/\mu_i)$ converges to 1 in probability as $N \to \infty$, and therefore we approximate $|\mathbb{T}_i(\zeta)|$ by $N\zeta/\mu_i$. To conclude; (12.130) follows from (12.15), (12.131), and (12.132). □

Proof of Theorem 12.1. Let $X_\zeta = Z_{\zeta/\mu_{\min}}$ denote the allele frequency process after changing time scale by a factor $\mu_{\min}$. Let $S_k = \mu_{\min}\tau_k$ refer to time points of fixation when $\{X_\zeta\}$ visits new fixed states in $\mathcal{Z}_{\text{hom}}$, defined in (12.6), $S'_{k+1} = \mu_{\min}\tau'_{k+1}$ the time point when a successful mutation first appears after S_k, and $S = \mu_{\min}T_m = S_M$ the time when allele m gets fixed. We need to show that

$$S \xrightarrow{\mathcal{L}} \text{PD}(\tilde{e}_0, \Sigma_0) \text{ as } N \to \infty. \tag{12.133}$$

To this end, write

$$S = \sum_{k=0}^{M-1}(S'_{k+1} - S_k) + \sum_{k=1}^{M}(S_k - S'_k) =: S_{\text{appear}} + S_{\text{tunfix}}, \tag{12.134}$$

where S_{appear} is the total waiting time for new successful mutations to appear, and S_{tunfix} is the total waiting time for tunneling and fixation, after successful mutations have appeared. We will first show that

$$S_{\text{appear}} \xrightarrow{\mathcal{L}} \text{PD}(\tilde{e}_0, \Sigma_0) \text{ as } N \to \infty. \tag{12.135}$$

It follows from (12.14) to (12.17) that $\{X_{S_k}\}$ is a Markov chain that starts at $X_{S_0} = e_0$, with transition probabilities

$$P(X_{S_{k+1}} = e_j | X_{S_k} = e_i) = p_{ij,N} \to \pi_{ij} \quad \text{for } i = 0, \ldots, m-1,\ j \neq i. \tag{12.136}$$

Because of (12.25) and Lemma 12.1, the waiting times for successful mutations $i \to i \pm 1$ have exponential or degenerate limit distributions as $N \to \infty$, since

$$P(S'_{k+1} - S_k > \zeta | X_{S_k} = e_i) \to \begin{cases} \exp(-\kappa_i \zeta), & i \in I_{\text{long}}, \\ 0, & i \in I_{\text{short}}, \end{cases} \tag{12.137}$$

where I_{long} and I_{short} refer to those asymptotic states in (12.22) and (12.23) that are visited for a long and short time, respectively. Since by definition, the non-asymptotic states $i \in I_{\text{nas}}$ in (12.20) will have no contribution to the limit distribution of S_{appear} as $N \to \infty$, it follows from (12.136) to (12.137) that asymptotically, S_{appear} is the total waiting time for a continuous time Markov chain with intensity matrix Σ, that starts at e_0, before it reaches its absorbing state e_m. This proves (12.135).

It remains to prove that S_{tunfix} is asymptotically negligible. It follows from (12.26) that

$$P(\varepsilon) = P_N(\varepsilon) = \max_{i \in I_{\text{as}}} P\left(S_k - S'_k > \varepsilon | X_{S_{k-1}} = e_i\right) = o(1) \tag{12.138}$$

as $N \to \infty$ for any $\varepsilon > 0$. Write $M = \sum_{i=0}^{m-1} M_i$, where M_i is the number of visits to e_i by the Markov chain $\{X_{S_k};\ k = 0, \ldots, M\}$, before it is stopped at time M. Let K be a large positive integer. We find that

$$\begin{aligned} P(S_{\text{tunfix}} > \varepsilon) &\leq E\left[\textstyle\sum_{k=1}^{\min(K,M)} P(S_k - S'_k > \varepsilon/K)\right] + P(M > K) \\ &\leq K P(\varepsilon/K) + \textstyle\sum_{i \in I_{\text{nas}}} P(M_i > 0) + E(M)/K \\ &\leq 2E(M)/K \end{aligned} \tag{12.139}$$

for all sufficiently large N. In the second step of (12.139) we used that

$$E(M) = \tilde{e}_0 (I - P_0)^{-1} 1^T \to \tilde{e}_0 (I - \Pi_0)^{-1} 1^T < \infty, \tag{12.140}$$

where P_0 is a square matrix of order m that contains the first m rows and m columns of the transition matrix P of the Markov chain X_{S_k}, so that its elements are the transition probabilities among and from the non-absorbing states. We used in (12.140) that M is the number of jumps until this Markov chain reaches its absorbing state, and therefore it has a discrete phase-type distribution (Bobbio et al. [9]). And because of (12.17)–(12.18), the expected value of M must be finite. In the last step of (12.139) we used (12.138) and the definition of non-asymptotic states, which implies $P(M_i > 0) = o(1)$ for all $i \in I_{\text{nas}}$.

Since (12.139) holds for all $K > 0$ and $\varepsilon > 0$, we deduce $S_{\text{tunfix}} = o(1)$ by first letting $K \to \infty$ and then $\varepsilon \to 0$. Together with (12.134)–(12.135) and Slutsky's Theorem (see for instance Gut [29]), this completes the proof of (12.133). □

In order to motivate Theorem 12.2, we first give four lemmas. It is assumed for all of them that the regularity conditions of Theorem 12.2 hold.

Lemma 12.2 *Let r_{ilj} be the probabilities defined in* (12.37)–(12.40). *Then*

$$\begin{aligned} r_{ilj} &= O(u_j^{1-2^{-(j-l-1)}}), \\ r_{ilj} &= \Omega(u_{l+2}^{1-2^{-(j-l-1)}}), \end{aligned} \quad i \le l \le j-2, \tag{12.141}$$

and

$$\begin{aligned} r_{ilj} &= O(v_j^{1-2^{-(l-j-1)}}), \\ r_{ilj} &= \Omega(v_{l-2}^{1-2^{-(l-j-1)}}), \end{aligned} \quad j+2 \le l \le i, \tag{12.142}$$

as $N \to \infty$. The corresponding formulas for $r_{ij} = r_{iij}$ in (12.36) *are obtained by putting $l = i$ in* (12.141)–(12.142).

Proof. In order to prove (12.141), assume $i \le l \le j-2$. Since $r_{i,j-1,j} = 1$, repeated application of the recursive formula $r_{i,k-1,j} = R(\rho_{ikj})\sqrt{r_{ikj}u_{k+1}}$ in (12.38), for $k = j-1, \ldots, l+1$, leads to

$$r_{ilj} = \prod_{k=l+1}^{j-1} R(\rho_{ikj})^{2^{-(k-l-1)}} u_{k+1}^{2^{-(k-l)}}. \tag{12.143}$$

We know from (12.48) that all $\rho_{ilj} = O(1)$ as $N \to \infty$. From this and the definition of the function $R(\rho)$ in (12.41), it follows that $R(\rho_{ilj}) = \Theta(1)$ as $N \to \infty$, so that

$$r_{ilj} = \Theta\left(\prod_{k=l+1}^{j-1} u_{k+1}^{2^{-(k-l)}}\right). \tag{12.144}$$

Then both parts of (12.141) follow by inserting the first equation of (12.46) into (12.144). The proof of (12.142) when $j+2 \le l \le i$ is analogous. Since $r_{i,j+1,j} = 1$, we use a recursion for $k = j+1, \ldots, l-1$ in order to arrive at the explicit formula

$$r_{ilj} = \prod_{k=j+1}^{l-1} R(\rho_{ikj})^{2^{-(l-k-1)}} v_{k-1}^{2^{-(l-k)}}.$$

Then use (12.48) and the third equation of (12.46) to verify that r_{ilj} satisfies (12.142). □

Lemma 12.3 *Let q_{ij}, q_{ilj}, r_{ij}, and r_{ilj} be the probabilities defined in connection with* (12.35)–(12.40). *Consider a fixed $i \in \{0, 1, \ldots, m-1\}$, and let $F(i)$ and $B(i)$ be the indices defined in* (12.44). *Then,*

$$\begin{aligned} &q_{ilF(i)} \sim r_{ilF(i)}, l = i, i+1, \ldots, F(i)-1, \\ &q_{ilB(i)} \sim r_{ilB(i)}, l = B(i)+1, \ldots, i, \text{ if } i > 0 \text{ and } \hat{\pi}_{iB(i)} > 0 \end{aligned} \tag{12.145}$$

as $N \to \infty$. *In particular,*

$$\begin{aligned} &q_{iF(i)} \sim r_{iF(i)}, \\ &q_{iB(i)} \sim r_{iB(i)}, \text{ if } i > 0 \text{ and } \hat{\pi}_{iB(i)} > 0. \end{aligned} \tag{12.146}$$

Sketch of proof. Notice that (12.146) is a direct consequence of (12.145), since $q_{iij} = q_{ij}$ and $r_{iij} = r_{ij}$. We will only motivate the upper part of (12.145), since the lower part is treated similarly. Consider a fixed $i \in \{0, \ldots, m-1\}$, and for simplicity of notation we write $j = F(i)$. We will argue that

$$q_{ilj} \sim r_{ilj} \tag{12.147}$$

for $l = j-1, \ldots, i$ by means of induction. Formula (12.147) clearly holds when $l = j-1$, since, by definition, $q_{i,j-1,j} = r_{i,j-1,j} = 1$. As for the induction step, let $i+1 \le l \le j-1$, and suppose (12.147) has been proved for l. Then recall the recursive formula

$$r_{i,l-1,j} = R(\rho_{ilj})\sqrt{u_{l+1} r_{ilj}} \tag{12.148}$$

from (12.38), with R defined in (12.41). If

$$q_{i,l-1,j} \sim R(\rho_{ilj})\sqrt{u_{l+1} q_{ilj}} \tag{12.149}$$

holds as well, then (12.147) has been shown for $l-1$, and the induction proof is completed. Without loss of generality we may assume that $j \ge i+2$, since otherwise the induction proof of (12.147) stops after the first trivial step $l = j-1$.

In order to motivate (12.149), we will look at what happens when the population is in fixed state i. Suppose $Z_{\tau_k} = e_i$, and recall that τ'_{k+1} is the time point when the first successful mutation $i \to i+1$ in (τ_k, τ_{k+1}) arrives. Therefore, if $Z_{\tau_{k+1}} = e_j$, there is a non-empty set $J = \{i+1, \ldots, j-1\}$ of types that must be present among some of the descendants of the successful mutation, before a mutation $j-1 \to j$ arrives at some time point $\tau''_{k+1} \in (\tau'_{k+1}, \tau_{k+1})$. Put $Z_{tJ} = \max_{l \in J} Z_{tl}$. The regularity condition

$$P\Big(\sup_{\tau'_{k+1} < t < \tau''_{k+1}} Z_{tJ} > \varepsilon | Z_{\tau_k} = e_i\Big) \to 0 \tag{12.150}$$

for all $\varepsilon > 0$ as $N \to \infty$, assures that with high probability, none of the alleles in J reaches a high frequency after the successful $i \to i+1$ mutation occurred, and before allele j first appears. We will need this condition below, for verifying the induction step (12.149).

The rationale for (12.150) is that fixation events $i \to j$ will happen much more frequently than other types of fixation events $i \to l$ with $l \in J$, because of (12.44). We will motivate that

$$P = P\Big(\sup_{\tau_k < t < \tau_k + a\mu_i^{-1}} Z_{tJ} > \varepsilon | Z_{\tau_k} = e_i\Big) \to 0 \tag{12.151}$$

for any $a > 0$ and $\varepsilon > 0$ as $N \to \infty$, with μ_i the rate of leaving fixation state i. In Lemma 12.1 we motivated that $\tau'_{k+1} - \tau_k = O_p(\mu_i^{-1})$, and in Lemma 12.5 we will argue that $\tau''_{k+1} - \tau'_{k+1} = o_p(\mu_i^{-1})$. Since this implies $\tau''_{k+1} - \tau_k = O_p(\mu_i^{-1})$, formula (12.150) will follow from (12.151).

In order to motivate (12.151), assume for simplicity there are no backward mutations (the proof is analogous but more complicated if we include back mutations as well). If allele $l \in J$ exceeds frequency ε, we refer to this as a semi-fixation event. Let $\lambda_{il}(\varepsilon)$ be the rate at which this happens after time τ_k, and before the next fixed state is reached. Then, the rate at which semi-fixation events happen among some $l \in J$, is

$$\begin{aligned}
\lambda_{iJ}(\varepsilon) &= \sum\nolimits_{l\in J} \lambda_{il}(\varepsilon) \\
&\sim N u_{i+1} \sum\nolimits_{l\in J} q_{il} \beta_{N\varepsilon}\left(\frac{s_l}{s_i}\right) \\
&\leq C(\varepsilon) \times N u_{i+1} \sum\nolimits_{l\in J} q_{il} \beta\left(\frac{s_l}{s_i}\right) \\
&\sim C(\varepsilon) \sum\nolimits_{l\in J} \lambda_{il}.
\end{aligned} \tag{12.152}$$

In the second step of (12.152) we introduced $\beta_{N\varepsilon}(s)$, the probability that a single mutant with fitness s reaches frequency ε, if all other individuals have fitness 1 and there are no mutations. We made use of

$$\lambda_{il}(\varepsilon) \sim N u_{i+1} q_{il} \beta_{N\varepsilon}\left(\frac{s_l}{s_i}\right). \tag{12.153}$$

This is motivated as in the proof of Lemma 12.4, in particular Eqs. (12.163), (12.164) and variant of (12.167) for semi-fixation rather than fixation. In the third step of (12.152) we utilized that $\beta_{N\varepsilon}(s)$ is larger than the corresponding fixation probability $\beta(s) = \beta_N(s)$ for a population of size N. In order to quantify how much larger the fixation probability of the smaller population of size $N\varepsilon$ is, we introduced $C(\varepsilon)$, an upper bound of $\beta_{N\varepsilon}(s_l/s_i)/\beta(s_l/s_i)$ that holds for all $l \in J$. An expression for $C(\varepsilon)$ can be derived from (12.32) if s_l/s_i is sufficiently close to 1. Indeed, we know from (12.48) that $s_l/s_i \to 1$ as $N \to \infty$. However, we need to sharpen this condition somewhat, to

$$s = \frac{s_l}{s_i} \geq 1 + \frac{x}{N} \tag{12.154}$$

for all $l \in J$ and some fixed $x < 0$. Then it follows from (12.32) that

$$\frac{\beta_{N\varepsilon}(s)}{\beta_N(s)} = \frac{s^{-N} - 1}{s^{-N\varepsilon} - 1} \leq \frac{(1 + x/N)^{-N} - 1}{(1 + x/N)^{-N\varepsilon} - 1} \to \frac{e^{-x} - 1}{e^{-\varepsilon x} - 1} =: C(\varepsilon)$$

is a constant not depending on N. Finally, in the last step of (12.152) we assumed

$$\lambda_{il} \sim N u_{i+1} q_{il} \beta \left(\frac{s_l}{s_i} \right), \quad l \in J. \tag{12.155}$$

This is motivated in the same way as Eq. (12.153), making use of (12.163)–(12.164) and (12.167).

Assuming that semi-fixation events arrive according to a Poisson process with intensity $\lambda_{iJ}(\varepsilon)$, formula (12.151) follows from (12.44) to (12.152), since

$$\begin{aligned} P &\sim 1 - \exp\left(-\lambda_{iJ}(\varepsilon) \times \frac{a}{\mu_i} \right) \\ &\le 1 - \exp\left(-C(\varepsilon) \sum\nolimits_{l \in J} \lambda_{il} \times \frac{a}{\mu_i} \right) \\ &= 1 - \exp(-C(\varepsilon) a \sum\nolimits_{l \in J} p_{il}) \\ &\to 1 - \exp(-C(\varepsilon) a \sum\nolimits_{l \in J} \pi_{il}) \\ &= 1 - \exp(-C(\varepsilon) a \sum\nolimits_{l \in J} \hat{\pi}_{il}) \\ &= 0 \end{aligned} \tag{12.156}$$

as $N \to \infty$. In the third step of (12.156) we used (12.16) to conclude that $p_{il} = \lambda_{il}/\mu_i$, and in the fourth step we utilized (12.17). In the fifth step of (12.156) we claimed that $\pi_{il} = \hat{\pi}_{il}$ for $l \in J$, Although we have not given a strict proof of this, it seems reasonable in view of the definitions of π_{il} and $\hat{\pi}_{il}$ in (12.17) and (12.43), together with (12.35), (12.155), and the fact that $q_{il} \sim r_{il}$ for $i < l < F(i)$ (which can be proved by induction with respect to l). Finally, in the last step of (12.156) we invoked (12.44), which implies $\hat{\pi}_{il} = 0$ for all $l \in J = \{i+1, \ldots, F(i)-1\}$.

Equation (12.150) enables us to approximate the allele frequency Z_{tl} by a branching process with mutations, in order to motivate (12.149). (A strict proof of this for a neutral model $s_0 = \cdots s_{m-1} = 1$ can be found in Theorem 2 of Durrett et al. [20].) We will look at the fate of the first $l-1 \to l$ mutation at time $\tau \in (\tau'_{k+1}, \tau''_{k+1})$, that is a descendant of the first successful $i \to i+1$ mutation at time τ'_{k+1}, and arrives before the first $j-1 \to j$ mutation at time τ''_{k+1}. Recall that $q = q_{i,l-1,j}$ is the probability that this l mutation gets an offspring that mutates into type j, and $q' = q_{ilj}$ is the corresponding probability that one of its descendants, an $l \to l+1$ mutation, gets a type j offspring. Let also $r' = r_{ilj}$ be the approximation q', and write $s = s_l/s_i$ for the ratio between the selection coefficients of alleles l and i. With this simplified notation, according to (12.149), we need to show that

$$q \sim R(\rho)\sqrt{uq'} \tag{12.157}$$

as $N \to \infty$, where $u = u_{l+1}$, and $\rho = \rho_{ilj}$ is defined in (12.37), i.e.

$$s = 1 + \rho\sqrt{ur'}. \tag{12.158}$$

We make the simplifying assumption that at time τ, the population has one single type l individual, the one that mutated from type $l-1$ at this time point, whereas all other $N-1$ individuals have type i. (Recall that we argued in Lemma 12.1 that such an assumption is asymptotically accurate.) In order to compute the probability q for the event A that this individual gets a descendant of type j, we condition on the next time point when one individual dies and is replaced by the offspring of an individual that reproduces. Let D and R be independent indicator variables for the events that the type l individual dies and reproduces respectively. Using the definition of the Moran process in Sect. 12.2, this gives an approximate recursive relation

$$\begin{aligned}
q &= P(A) \\
&= P(D=0, R=0)P(A|D=0, R=0) \\
&+ P(D=0, R=1)P(A|D=0, R=1) \\
&+ P(D=1, R=0)P(A|D=1, R=0) \\
&+ P(D=1, R=1)P(A|D=1, R=1) \\
&= \left(1-\frac{1}{N}\right)\frac{N-1}{N-1+s} \times q \\
&+ \left(1-\frac{1}{N}\right)\frac{s}{N-1+s} \\
&\times \left[u(q'+q-q'q) + vq + (1-u-v)(2q-q^2)\right] \\
&+ \frac{1}{N}\frac{N-1}{N-1+s} \times 0 \\
&+ \frac{1}{N}\frac{s}{N-1+s} \times \left[uq' + v \times 0 + (1-u-v)q\right]
\end{aligned} \tag{12.159}$$

for q, where $v = v_{l-1}$ is the probability of a back mutation $l \to l-1$. In the last step of (12.159) we retained the exact transition probabilities of the Moran process, but we used a branching process approximation for the probability q that the type l mutation at time τ gets a type j descendant. This approximation relies on (12.150), and it means that descendants of the type l mutation that are alive at the same time point, have independent lines of descent after this time point. For instance, in the second term on the right hand side of (12.159), a type i individual dies and the type l individual reproduces ($D=0$, $R=1$). Then there are three possibilities: First, the offspring of the type l individual mutates to $l+1$ with probability u. Since the type l individual and its type $l+1$ offspring have independent lines of descent, the probability is $1-(1-q')(1-q) = q'+q-q'q$ that at least one of them gets a type j descendant. Second, if the offspring mutates back to $l-1$ (with probability v), its type l parent has a probability q of getting a type j descendant. Third, if the offspring does not mutate (with probability $1-u-v$), there are two type l individuals, with a probability $1-(1-q)^2 = 2q-q^2$ that at least one of them gets a type j offspring.

Equation (12.159) is quadratic in q. Dividing both sides of it by $s/(N-1+s)$, it can be seen, after some computations, that this equation simplifies to $aq^2+bq+c=0$, with

$$\begin{aligned} a &= (1-u-v)\left(1-\frac{1}{N}\right) \sim 1, \\ b &= \frac{N-1}{N}\times\frac{1-s}{s}+u(1+q'-\frac{q'}{N})+v \\ &\sim -\frac{\rho\sqrt{ur'}}{1+\rho\sqrt{ur'}}+(1+q')u+v \\ &\sim -\rho\sqrt{uq'}, \\ c &= -uq', \end{aligned} \tag{12.160}$$

as $N\to\infty$. When simplifying the formula for b, we used (12.158) in the second step, the induction hypothesis (12.147) in the last step (since it implies $q'\sim r'$), and additionally we assumed in the last step that $(1+q')u+v=o(\sqrt{ur'})$. In order to justify this, from the second equation of (12.46) we know that $v=O(u)$, and since $q'\le 1$, it suffices to verify that $u=o(\sqrt{ur'})$, or equivalently that $r'=\Omega(u)$. But this follows from (12.46), (12.141), and the fact that $u=u_{l+1}$, since

$$r'=r_{ilj}=\Omega\left(u_{l+2}^{1-2^{-(j-l-1)}}\right)=\Omega\left(u^{1-2^{-(j-l-1)}}\right)=\Omega(u),$$

where in the last step we used that $l\le j-1$. This verifies the asymptotic approximation of b in (12.160).

To conclude, in order to prove of (12.157), we notice that the only positive solution to the quadratic equation in q, with coefficients as in (12.160), is

$$\begin{aligned} q &\sim \frac{\rho\sqrt{uq'}}{2}+\sqrt{\frac{\rho^2uq'}{4}+uq'} \\ &= \frac{\rho+\sqrt{\rho^2+4}}{2}\sqrt{uq'} \\ &= R(\rho)\sqrt{uq'}, \end{aligned}$$

where in the last step we invoked the definition of $R(\rho)$ in (12.41). This finishes the proof of the induction step (12.149) or (12.157), and thereby the proof of (12.147).

We end this proof by a remark: Recall that r_{ij} in (12.36) is an approximation q_{ij}, obtained from recursion (12.38) or (12.148) when $j>i$, and from (12.40) when $j<i$. A more accurate (but less explicit) approximation of q_{ij} is obtained, when $i<j$, by recursively solving the quadratic equation $ax^2+bx+c=0$, with respect to $x=r_{i,l-1,j}$ for $l=j-1,\dots,i+1$, and finally putting $r_{ij}=r_{iij}$. The coefficients of this equation are defined as in (12.160), with $r'=r_{ilj}$ instead of q'. When $j<i$, the improved approximation of q_{ij} is defined analogously. □

Lemma 12.4 *Let μ_i be the rate (12.15) at which a successful forward or backward mutation occurs in a homogeneous type i population, and let $\hat{\mu}_i$ in (12.42) be its approximation. Define the asymptotic transition probabilities π_{ij} between fixed population states as in (12.17), and their approximations $\hat{\pi}_{ij}$ as in (12.43). Then*

$$\mu_i \sim \hat{\mu}_i, \quad i = 0, \ldots, m-1, \tag{12.161}$$

as $N \to \infty$, and

$$\pi_{ij} = \hat{\pi}_{ij}, \quad i, j = 0, 1, \ldots, m. \tag{12.162}$$

Sketch of proof. Consider a time point τ_k when the population becomes fixed with type i, so that $Z_{\tau_k} = e_i$. Denote by f_{ij} the probability a forward mutation $i \to i+1$, which appears at a time point later than τ_k, is the first successful mutation after τ_k, that its descendants have taken over the population by time τ_{k+1}, and that all of them by that time have type j (so that $Z_{\tau_{k+1}} = e_j$). Likewise, when $j < i$ and $i \geq 1$, we let b_{ij} refer to the probability that if a backward mutation $i \to i-1$ arrives, it is successful, its descendants have taken over the population by time τ_{k+1}, and all of them have type j. For definiteness we also put $b_{0j} = 0$. We argue that

$$\lambda_{ij} \sim \begin{cases} N u_{i+1} f_{ij}, & j > i, \\ N v_{i-1} b_{ij}, & j < i, \end{cases} \tag{12.163}$$

since the event that the population at time τ_{k+1} have descended from more than one $i \to i \pm 1$ mutation that occurred in the time interval (τ_k, τ_{k+1}), is asymptotically negligible.

Let $\beta_j(z)$ be the probability that the descendants of a type j individual, who lives in a population with a type configuration z, takes over the population so that it becomes homogeneous of type j. Although $\beta_j(z)$ depends on the mutation rates $u_1, \ldots, u_m, v_0, \ldots, v_{m-1}$ as well as the selection coefficients $s_1, \ldots, s_m$, this is not made explicit in the notation. The probabilities f_{ij} and b_{ij} in (12.163) can be written as a product

$$\begin{aligned} f_{ij} &= q_{ij} E\left[\beta_j(Z_{\tau''_{k+1}}) | A_j, Z_{\tau_k} = e_i\right], \; j > i, \\ b_{ij} &= q_{ij} E\left[\beta_j(Z_{\tau''_{k+1}}) | A_j, Z_{\tau_k} = e_i\right], \; j < i \end{aligned} \tag{12.164}$$

of two terms. Recall that the first term, q_{ij}, is the probability that the first successful mutation $i \to i \pm 1$ at time $\tau'_{k+1} > \tau_k$ has a descendant that mutates into type j at some time $\tau''_{k+1} \in (\tau'_{k+1}, \tau_{k+1})$. The second term is the probability that this mutation has spread to the rest of the population by time τ_{k+1}. The conditional expectation of this second term is with respect to variations in $Z_{\tau''_{k+1}}$, and the conditioning is with respect to A_j, the event that the mutation at time τ''_{k+1} is into type j.

In order to compare the transition rates in (12.163) with the approximate ones in (12.35), we notice that the latter can be written as

$$\hat{\lambda}_{ij} = \begin{cases} Nu_{i+1}\hat{f}_{ij}, & j > i, \\ Nv_{i-1}\hat{b}_{ij}, & j < i, \end{cases} \tag{12.165}$$

where

$$\begin{aligned} \hat{f}_{ij} &= r_{ij}\beta(s_j/s_i), \quad j > i, \\ \hat{b}_{ij} &= r_{ij}\beta(s_j/s_i), \quad j < i, \end{aligned} \tag{12.166}$$

r_{ij} is the approximation of q_{ij} defined in (12.36), whereas $\beta(s_j/s_i)$ is the probability that a single type j individual gets fixed in a population without mutations, where all other individuals have type i.

We will argue that the probabilities in (12.166) are asymptotically accurate approximations of those in (12.164), for all pairs i, j of states that dominate asymptotically, that is, those pairs for which $j \in \{B(i), F(i)\}$. In Lemma 12.3 we motivated that r_{ij} is an asymptotically accurate approximation of q_{ij} for all such pairs of states. Likewise, we argue that $\beta(s_j/s_i)$ is a good approximation of the conditional expectation in (12.164). Indeed, following the reasoning of Lemma 12.3, since none of the intermediate alleles, between i and j, will reach a high frequency before the type j mutant appears at time τ''_{k+1}, it follows that most of the other $N-1$ individuals will have type i at this time point. Consequently,

$$E\left[\beta_j(Z_{\tau''_{k+1}})|A_j, Z_{\tau_k} = e_i\right] \sim \beta_j\left(\frac{N-1}{N}e_i + \frac{1}{N}e_j\right) \sim \beta\left(\frac{s_j}{s_i}\right) \tag{12.167}$$

as $N \to \infty$. In the last step of (12.167) we used that new mutations between time points τ''_{k+1} and τ_{k+1} can be ignored, because of the smallness (12.4)–(12.5) of the mutation rates. Since $\beta_j\left((N-1)e_i/N + e_j/N\right)$ is the fixation probability of a single type j mutant that has selection coefficient s_j/s_i relative to the other $N-1$ type i individuals, it is approximately equal to the corresponding fixation probability $\beta(s_j/s_i)$ of a mutation free Moran model. It therefore follows from (12.164) and (12.166) that

$$\begin{aligned} \hat{f}_{iF(i)} &\sim f_{iF(i)}, i = 0, \ldots, m-1, \\ \hat{b}_{iB(i)} &\sim b_{iB(i)}, i = 1, \ldots, m-1 \text{ and } B(i) \neq \emptyset \end{aligned} \tag{12.168}$$

as $N \to \infty$.

Next we consider pairs of types i, j such that $j \notin \{B(i), F(i)\}$. We know from (12.44), (12.165) and (12.166) that $\hat{f}_{il} = o(\hat{f}_{iF(i)})$ for all $l > i$ such that $l \neq F(i)$. It is therefore reasonable to assume that $f_{il} = o(f_{iF(i)})$ as well for all $l > i$ with $l \neq F(i)$, although $\hat{f}_{il}$ need not necessarily be a good approximation of f_{il} for all these l. The same argument also applies to backward mutations when $B(i) \neq \emptyset$ and $\hat{\pi}_{iB(i)} > 0$, that is, we should have $f_{il} = o(f_{iB(i)})$ for all $l < i$ such that $l \neq B(i)$.

Putting things together, it follows from (12.44), (12.163), (12.165), (12.168), and the last paragraph that the approximate rate (12.42) at which a homogeneous type i population is transferred into a new fixed state, satisfies

$$\begin{aligned}\hat{\mu}_i &= Nv_{i-1}\sum_{j=0}^{i-1}\hat{b}_{ij} + Nu_{i+1}\sum_{j=i+1}^{m}\hat{f}_{ij}\\ &\sim 1\left(\hat{\pi}_{iB(i)} > 0\right) Nv_{i-1}\hat{b}_{iB(i)} + Nu_{i+1}\hat{f}_{iF(i)}\\ &\sim 1\left(\hat{\pi}_{iB(i)} > 0\right) Nv_{i-1}b_{iB(i)} + Nu_{i+1}f_{iF(i)}\\ &\sim Nv_{i-1}\sum_{j=0}^{i-1}b_{ij} + Nu_{i+1}\sum_{j=i+1}^{m}f_{ij}\\ &\sim \mu_i,\end{aligned} \tag{12.169}$$

as $N \to \infty$, in agreement with (12.161). Formulas (12.16)–(12.17), (12.43)–(12.44), (12.163), (12.165), and (12.168)–(12.169) also motivate why π_{ij} should equal $\hat{\pi}_{ij}$, in accordance with (12.162). □

Lemma 12.5 *The regularity condition* (12.47) *of Theorem 12.2 implies that* (12.26) *holds.*

Sketch of proof. Suppose $Z_{\tau_k} = e_i$ and $Z_{\tau_{k+1}} = e_j$ for some $i \in I_{\text{as}}$ and $j \neq i$. Write

$$\tau_{k+1} - \tau'_{k+1} = \begin{cases} \sum_{l=i+1}^{j-1}\sigma_l + \sigma_{\text{fix}} := \sigma_{\text{tunnel}} + \sigma_{\text{fix}}, & j > i,\\ \sum_{l=j+1}^{i-1}\sigma_l + \sigma_{\text{fix}} := \sigma_{\text{tunnel}} + \sigma_{\text{fix}}, & j < i.\end{cases} \tag{12.170}$$

If $j > i$, then the successful mutation at time τ'_{k+1} is from i to $i+1$. This type $i+1$ mutation has a line of descent with individuals that mutate to types $i+2, \ldots, j$, before the descendants of the type j mutation take over the population. The first term $\sigma_{\text{tunnel}} = \tau''_{k+1} - \tau'_{k+1}$ on the right hand side of (12.170) is the time it takes for the type $i+1$ mutation to tunnel into type j. It is the sum of σ_l, the time it takes for the type $l+1$ mutation to appear after the type l mutation, for all $l = i+1, \ldots, j-1$. The second term $\sigma_{\text{fix}} = \tau_{k+1} - \tau''_{k+1}$ on the right hand side of (12.170) is the time it takes for j to get fixed after the j mutation first appears. When $j < i$, we interpret the terms of (12.170) analogously. It follows from (12.170) that in order to prove (12.26), it suffices to show that

$$\begin{aligned}\sigma_{\text{tunnel}} &= o_p(\mu_{\min}^{-1}),\\ \sigma_{\text{fix}} &= o_p(\mu_{\min}^{-1}),\end{aligned} \tag{12.171}$$

as $N \to \infty$ for all asymptotic states $i \in I_{\text{as}}$. When $j > i$, we know from (12.44) to (12.162) that with probability tending to 1, $j = F(i)$. Following the argument from the proof of Theorem 2 of Durrett et al. [20], we have that

$$\sigma_l = O_p(q_{ilj}^{-1}). \tag{12.172}$$

In the special case when $l = i+1$ and $j = i+2$, formula (12.172) can also be deduced from the proof of Theorem 12.3, by looking at $G(x)/G(\infty)$ in (12.191). Using (12.172), we obtain the upper part of (12.171), since

$$\begin{aligned}\sigma_{\text{tunnel}} &= \sum_{l=i+1}^{j-1} \sigma_l \\ &= O_p\left(\sum_{l=i+1}^{j-1} q_{ilj}^{-1}\right) \\ &= o_p(q_{iij}^{-1}) \\ &= o_p(q_{ij}^{-1}) \\ &= o_p(\mu_i^{-1}) \\ &= o_p(\mu_{\min}^{-1}).\end{aligned} \tag{12.173}$$

In the second step of (12.173) we used that $q_{iij} \leq q_{ilj}$ for $i < l$, which follows from the definition of these quantities, in the third step we invoked $q_{ij} = q_{iij}$, and in the fourth step we applied the relation

$$\mu_i = \Theta\left(N u_{i+1} q_{ij} \beta\left(\frac{s_i}{s_j}\right)\right) = o(q_{ij}). \tag{12.174}$$

The first step of (12.174) is motivated as in Lemma 12.4, since $j = F(i)$ and hence $\pi_{ij} > 0$, whereas the second step follows from (12.4) and the fact that $\beta(s_i/s_j)$ is bounded by 1. Finally, the fourth step of (12.173) follows from the definition of $\mu_{\min}$ in (12.24), since (12.174) applies to any $i \in I_{\text{as}}$. When $j < i$, the first part of (12.171) is shown analogously.

In order to verify the second part of (12.171), we know from the motivation of Lemma 12.4 that with high probability, σ_{fix} is the time it takes for descendants of the type j mutation to take over the population, ignoring the probability that descendants of other individuals first mutated into j and then some of them survived up to time τ_{k+1} as well. We further recall from Lemma 12.4 that because of the smallness (12.4)–(12.5) of the mutation rates, right after the j mutation has arrived at time τ''_{k+1}, we may assume that the remaining $N-1$ individuals have type i, and after that no other mutation occurs until the j allele gets fixed at time τ_{k+1}. With these assumptions, σ_{fix} is the time for one single individual with selection coefficient s_j/s_i to get fixed in a two-type Moran model without mutations, where all other individuals have selection coefficient 1. From Sect. 12.5 it follows that $E(\sigma_{\text{fix}}) \sim \alpha(s_j/s_i)$, and therefore the second part of (12.171) will be proved if we can verify that

$$\alpha\left(\frac{s_j}{s_i}\right) = o(\mu_{\min}^{-1})$$

holds for all $i \in I_{\text{as}}$ and $j \in \{B(i), F(i)\}$ as $N \to \infty$. This is equivalent to showing that

$$\mu_{\min} = o\left(\min_{i \in I_{\text{as}}} \min\left[\alpha^{-1}\left(\frac{s_{B(i)}}{s_i}\right), \alpha^{-1}\left(\frac{s_{F(i)}}{s_i}\right)\right]\right) \tag{12.175}$$

as $N \to \infty$, where the $\alpha^{-1}(s_{B(i)}/s_i)$-term is included only when $B(i) \neq \emptyset$ (or equivalently, when $\pi_{iB(i)} > 0$). Using (12.44), (12.46), (12.141), (12.161), (12.168), and

(12.169), we find that

$$\begin{aligned}\mu_i \sim & \hat{\mu}_i \\ =& O\left(N u_{i+1} r_{iF(i)} \beta(s_{F(i)}/s_j)\right) \\ =& O\left(N u_{i+1} u_{F(i)}^{1-2^{-(F(i)-i-1)}} \beta(s_{F(i)}/s_j)\right) \\ =& O\left(N u_{F(i)}^{2-2^{-(F(i)-i-1)}} \beta(s_{F(i)}/s_j)\right).\end{aligned} \tag{12.176}$$

Inserting (12.176) into the definition of $\mu_{\min}$ in (12.24), we obtain

$$\mu_{\min} = O\left(\min_{i \in I_{\text{long}}} N u_{F(i)}^{2-2^{-(F(i)-i-1)}} \beta(s_{F(i)}/s_j)\right),$$

and formula (12.175) follows, because of (12.47). □

Proof of Theorem 12.2. We need to establish that the limit result (12.49) of Theorem 12.2 follows from Theorem 12.1. To this end, we first need to show that all $\hat{\lambda}_{ij}$ are good approximations of λ_{ij}, in the sense specified by Theorem 12.2, i.e. $\pi_{ij} = \hat{\pi}_{ij}$ and $\hat{\mu}_i / \hat{\mu}_{\min} \to \kappa_i$ as $N \to \infty$. But this follows from Lemma 12.4, and the definitions of $\mu_{\min}$ and $\hat{\mu}_{\min}$ in (12.24) and Theorem 12.2. Then it remains to check those two regularity conditions (12.18) and (12.26) of Theorem 12.1 that are not present in Theorem 12.2. But (12.18) follows from (12.44) to (12.162), since these two equations imply $\pi_{iF(i)} > 0$ for all $i = 0, \ldots, m-1$, and (12.26) follows from Lemma 12.5. □

Proof of (12.109). Let

$$\theta_i = u \times E(T_m | Z_0 = e_i) \tag{12.177}$$

be the standardized expected waiting time until all m mutations have appeared and spread in the population, given that it starts in fixed state i. Our goal is to find an explicit formula for θ_0, and then show that (12.109) is an asymptotically accurate approximation of this explicit formula as $m \to \infty$.

Recall that Σ_{ij} in (12.107) are the elements of the intensity matrix, for the Markov process that switches between fixed population states, when time has been multiplied by $\hat{\mu}_{\min} = u$. When the population is in fixed state i, the standardized expected waiting time until the next transition is $1/(-\Sigma_{ii})$. By conditioning on what happens at this transition, it can be seen that the standardized expected waiting times in (12.177), satisfy a recursive relation

$$\theta_i = \frac{1}{-\Sigma_{ii}} + \frac{\Sigma_{i,i-1}}{-\Sigma_{ii}} \times \theta_{i-1} + \frac{\Sigma_{i,i+1}}{-\Sigma_{ii}} \times \theta_{i+1}, \tag{12.178}$$

for $i = 0, 1, \ldots, m-1$, assuming $\theta_{-1} = 0$ on the right hand side of (12.178) when $i = 0$, and similarly $\theta_m = 0$ when $i = m-1$. Inserting the values of Σ_{ij} from (12.107) into (12.178), we can rewrite the latter equation as

$$\theta_0 - \theta_1 = \frac{1}{m} =: b_0 \tag{12.179}$$

and

$$\theta_i - \theta_{i+1} = \frac{Ci}{m-i}(\theta_{i-1} - \theta_i) + \frac{1}{m-i} =: a_i(\theta_{i-1} - \theta_i) + b_i, \tag{12.180}$$

for $i = 1, \ldots, m-1$, respectively. We obtain an explicit formula for θ_0 by first solving the linear recursion for $\theta_i - \theta_{i+1}$ in (12.179)–(12.180), and then summing over i. This yields

$$\theta_0 = \sum_{i=0}^{m-1}(\theta_i - \theta_{i+1}) = \sum_{i=0}^{m-1}\sum_{k=0}^{i}\theta_{ik}, \tag{12.181}$$

where

$$\theta_{ik} = b_k \prod_{j=k+1}^{i} a_j = \frac{\binom{m-1}{k}}{(m-k)\binom{m-1}{i}} \times C^{i-k}. \tag{12.182}$$

Formulas (12.181)–(12.182) provide the desired explicit formula for θ_0. When $C = 0$, it is clear that

$$\begin{aligned}\theta_0 &= \sum_{i=0}^{m-1}\theta_{ii}\\ &= \sum_{i=0}^{m-1} 1/(m-i)\\ &\sim \log(m) + \gamma,\end{aligned}$$

where $\gamma \approx 0.5772$ is the Euler–Mascheroni constant. This proves the upper half of (12.109). For $C > 0$, we will show that when m gets large, the (standardized) expected waiting time until the last mutant gets fixed, $\theta_{m-1} - \theta_m = \theta_{m-1}$, dominates the first sum in (12.181). To this end, we first look at θ_{m-1}, and rewrite this quantity as

$$\begin{aligned}\theta_{m-1} &= \sum_{k=0}^{m-1}\theta_{m-1,k}\\ &= \tfrac{1}{\binom{m-1}{m-1}}\sum_{k=0}^{m-1}\tfrac{1}{m-k}\binom{m-1}{k}C^{m-1-k}\\ &= (1+C)^{m-1}\sum_{k=0}^{m-1}\tfrac{1}{m-k}\binom{m-1}{k}\left(\tfrac{1}{1+C}\right)^k\left(\tfrac{C}{1+C}\right)^{m-1-k}\\ &= (1+C)^{m-1}E\left(\tfrac{1}{m-X_{m-1}}\right)\\ &= (1+C)^{m-1}E\left(\tfrac{1}{1+Y_{m-1}}\right),\end{aligned} \tag{12.183}$$

where

$$\begin{aligned}X_{m-1} &\overset{\mathcal{L}}{\in} \mathrm{Bin}\left(m-1, \tfrac{1}{1+C}\right),\\ Y_{m-1} = m-1-X_{m-1} &\overset{\mathcal{L}}{\in} \mathrm{Bin}\left(m-1, \tfrac{C}{1+C}\right)\end{aligned}$$

are two binomially distributed random variables. For large m, we apply the Law of Large Numbers to Y_{m-1} and find that

$$\begin{aligned}\theta_{m-1} \approx &(1+C)^{m-1}\tfrac{1}{1+E(Y_{m-1})}\\ \approx &(1+C)^{m-1}\tfrac{1}{mC/(1+C)}\\ = &(1+C)^m/(Cm),\end{aligned} \tag{12.184}$$

in agreement with the lower half of (12.109). In view of (12.181), in order to finalize the proof of (12.109), we need to show that the sum of $\theta_{m-j}-\theta_{m-j+1}$ for $j=2,3,\ldots,m$, is of a smaller order than (12.184). A similar argument as in (12.183) leads to

$$\begin{aligned}\theta_{m-j}-\theta_{m-j+1} = &\sum\nolimits_{k=0}^{m-j}\theta_{m-j,k}\\ = &(j-1)!(1+C)^{m-j}E\left[\tfrac{1}{\prod_{n=1}^{j}(n+Y_{m-j})}\right]\\ \le &\tfrac{2}{j}(1+C)^{m-j}E\left[\tfrac{1}{(1+Y_{m-j})(2+Y_{m-j})}\right],\end{aligned} \tag{12.185}$$

where

$$Y_{m-j} \overset{\mathcal{L}}{\in} \mathrm{Bin}\left(m-j,\frac{C}{1+C}\right).$$

For large m we have, by the Law of Large Numbers, that

$$\begin{aligned}&\theta_{m-j}-\theta_{m-j+1} \le \tfrac{2}{j}(1+C)^{m-j}\tfrac{1}{[1+(m-j)C/(1+C)]^2}\\ \le &\begin{cases}4(1+C)^{m/2}/m, & j>m/2,\\ (1+C)^{m-j}/\left[m/2\times C/(1+C)\right]^2, & 2\le j\le m/2.\end{cases}\end{aligned} \tag{12.186}$$

By summing (12.186) over j, it is easy to see that

$$\sum_{j=2}^{m}(\theta_{m-j}-\theta_{m-j+1}) \ll (1+C)^m/(Cm) \sim \theta_m$$

as $m\to\infty$. Together with (12.184), this completes the derivation of the lower part of (12.109). □

Sketch of proof of Theorem 12.3. Our proof will parallel that of Theorem 1 in Durrett el al. [20], see also Wodarz and Komarova [66]. We first use formula (12.66) in order to deduce that the ratio between the two rates of fixation from a type 0 population, satisfies $\hat{\lambda}_{02}/\hat{\lambda}_{01}\to\infty$ as $N\to\infty$. When $\rho=0$ in (12.51), this is a consequence of $\hat{\lambda}_{02}/\hat{\lambda}_{01}\sim N\sqrt{u_2}$ and the assumption $N\sqrt{u_2}\to\infty$ on the second mutation rate u_2. When $\rho<0$, $\hat{\lambda}_{02}/\hat{\lambda}_{01}$ tends to infinity at an even faster rate, due to the $\psi(\rho u_2^{1/2})$-term

of $\hat{\lambda}_{01}$ in (12.66). In any case, it follows that condition (12.44) is satisfied, with $F(0) = 2$ and $\hat{\pi}_{02} = 1$. That is, tunneling from 0 to 2 will occur with probability tending to 1 as $N \to \infty$ whether $\rho = 0$ or $\rho < 0$. As in the proof of Lemma 12.3 we conclude from this that the fraction $Z_t = Z_{t1}$ of allele 1 will stay close to 0, and we may use a branching process approximation for Z_t. A consequence of this approximation is that type 1 mutations arrive according to a Poisson process with intensity Nu_1, and the descendants of different type 1 mutations evolve independently. Let $0 < \sigma \leq \infty$ be the time it takes for the first type 2 descendant of a type 1 mutation to appear. In particular, if $\sigma = \infty$, this type 1 mutation has no type 2 descendants. Letting $G(x) = P(\sigma \leq x)$ be the distribution function of σ, it follows by a Poisson process thinning argument that

$$P(T_2'' \geq t) \sim \exp(-Nu_1 \int_0^t G(x)dx). \tag{12.187}$$

We use Kolmogorov's backward equation in order to determine G. To this end, we will first compute $G(x + h)$ for a small number $h > 0$, by conditioning on what happens during the time interval $(0, h)$. As in formulas (12.121)–(12.122) of Appendix B, we let $a_{ij}(z)$ refer to the rate at which a type i individual dies and gets replaced by the offspring of a type j individual, when the number of type 1 individuals before the replacement is Nz. Since we look at the descendants of one type 1 individual, we have that $z = Z_0 = 1/N$. Using a similar argument as in Eq. (12.159), it follows from this that

$$\begin{aligned} &G(x+h) = a_{00}(1/N)h \times G(x) \\ &+a_{01}(1/N)h\left[u_2 \times 1 + (1-u_2)(2G(x) - G(x)^2)\right] \\ &+a_{10}(1/N)h \times 0 + a_{11}(1/N)h \times [u_2 \times 1 + (1-u_2)G(x)] \\ &+\left[1 - \sum\nolimits_{ij} a_{ij}(1/N)h\right] G(x) + o(h) \end{aligned} \tag{12.188}$$

for small $h > 0$. Notice that the two $a_{00}(1/N)$ terms cancel out in (12.188), whereas $a_{11}(1/N)(1 - G(x))u_2 \times h = O(N^{-2}u_2 \times h)$ is too small to have an asymptotic impact. Using formulas (12.121)–(12.122) for $a_{01}(1/N)$ and $a_{10}(1/N)$, it follows that (12.188) simplifies to

$$G(x+h) = s \times h\left[u_2 + 2G(x) - G(x)^2\right] + 1 \times h \times 0 + [1 - (s+1)h]\, G(x) + o(h),$$

when all asymptotically negligible terms are put into the remainder term. Letting $h \to 0$, we find that $G(x)$ satisfies the differential equation

$$\begin{aligned} G'(x) &= -sG(x)^2 + (s-1)G(x) + su_2 \\ &= -s(G(x) - r_1)(G(x) - r_2), \end{aligned} \tag{12.189}$$

where

$$r_1 = (s-1)/(2s) + \sqrt{[(s-1)/(2s)]^2 + u_2},$$
$$r_2 = (s-1)/(2s) - \sqrt{[(s-1)/(2s)]^2 + u_2}$$

are the two roots of the quadratic equation $-sy^2 + (s-1)y + su_2 = 0$. Recall from (12.51) that $s = 1 + \rho\sqrt{u_2}$. We may therefore express these two roots as

$$\begin{aligned} r_1 &= \sqrt{u_2}\left(\rho + \sqrt{\rho^2 + 4s^2}\right)/(2s) \sim \sqrt{u_2}\left(\rho + \sqrt{\rho^2+4}\right)/(2s) \\ &= \sqrt{u_2}R(\rho)/s, \\ r_2 &= \sqrt{u_2}\left(\rho - \sqrt{\rho^2 + 4s^2}\right)/(2s) \sim \sqrt{u_2}\left(\rho - \sqrt{\rho^2+4}\right)/(2s), \end{aligned} \tag{12.190}$$

where in the second step we used that $u_2 \to 0$ and $s \to 1$ as $N \to \infty$, and in the last step we invoked (12.41), the definition of $R(\rho)$. Since $r_2 < 0 < r_1$, and $G'(x) \to 0$ as $x \to \infty$, it follows from (12.189) that we must have $G(\infty) = r_1$. Together with the other boundary condition $G(0) = 0$, this gives as solution

$$G(x) = r_1 \frac{1 - e^{-(r_1 - r_2)sx}}{1 - \frac{r_1}{r_2}e^{-(r_1-r_2)sx}} \tag{12.191}$$

to the differential equation (12.189), with

$$r_1 - r_2 \sim \frac{\sqrt{u_2} \times \sqrt{\rho^2+4}}{s}$$

and

$$-\frac{r_1}{r_2} \sim \frac{\sqrt{\rho^2+4} + \rho}{\sqrt{\rho^2+4} - \rho}. \tag{12.192}$$

Putting things together, we find that

$$\begin{aligned} P\left(NR(\rho)u_1\sqrt{u_2} \times T_2'' \geq t\right) &\sim P\left(Nu_1r_1s \times T_2'' \geq t\right)) \\ &\sim \exp\left(-Nu_1 \int_0^{t/(Nu_1r_1s)} G(x)dx\right) \\ &\sim \exp\left(-\int_0^t h(y)dy\right), \end{aligned} \tag{12.193}$$

where formula (12.190) was used in the first step, (12.187) in the second step, in the third step we changed variables $y = Nu_1r_1s \times x$ and introduced the hazard function $h(x) = G\left(x/(Nu_1r_1s)\right)/(sr_1)$. If $Nu_1 \to a > 0$ as $N \to \infty$, it follows from (12.191) and the fact that $s \to 1$ that we can rewrite the hazard function as

$$\begin{aligned} h(x) &\sim \tfrac{1}{sr_1} G\left(\tfrac{x}{sar_1}\right) \\ &= \tfrac{1}{s} \times \frac{1-\exp\left(-\frac{r_1-r_2}{r_1}\times\frac{x}{a}\right)}{1-\frac{r_1}{r_2}\exp\left(-\frac{r_1-r_2}{r_1}\times\frac{x}{a}\right)} \sim \frac{1-\exp\left(-\frac{r_1-r_2}{r_1}\times\frac{x}{a}\right)}{1-\frac{r_1}{r_2}\exp\left(-\frac{r_1-r_2}{r_1}\times\frac{x}{a}\right)}. \end{aligned} \tag{12.194}$$

We finally obtain the limit result (12.110)–(12.111) when $a > 0$ from (12.193) to (12.194), using (12.192) and the fact that

$$\frac{r_1 - r_2}{r_1} \sim \frac{2\sqrt{\rho^2+4}}{\rho + \sqrt{\rho^2+4}}.$$

When $Nu_1 \to 0$, one similarly shows that (12.193) holds, with $h(x) = 1$. Finally, formula (12.112) follows by integrating (12.193) with respect to t. □

Motivation of formula (12.114). We will motivate formula (12.114) in terms of the transition rates $\hat{\lambda}_{ij}$ in (12.35), rather than those in (12.113) that are adjusted for tunneling and fixation of alleles.

Since we assume $s_1 = \cdots = s_{m-1} = 1 < s_m$ in (12.114), it follows from (12.35) that it is increasingly difficult to have backward and forward transitions over larger distances, except that it is possible for some models to have a direct forward transition to the target allele m. By this we mean that the backward and forward transition rates from any state i satisfy $\hat{\lambda}_{i,i-1} \gg \cdots \gg \hat{\lambda}_{i0}$, and $\hat{\lambda}_{i,i+1} \gg \cdots \gg \hat{\lambda}_{i,m-1}$ respectively, as $N \to \infty$. For this reason, from any fixed state i, it is only possible to have competition between the two forward transitions $i \to i+1$ and $i \to m$ when $0 \le i \le m-2$. Since $\gamma_i = (\hat{\lambda}_{im}/\hat{\lambda}_{i,i+1})^2$, and since the transition rates to the intermediate alleles $i+1, \ldots, m-1$ are of a smaller order than the transition rate to $i+1$, it follows that (12.35) predicts a total forward rate of fixation from fixed state i of the order

$$\begin{aligned} Nu_{i+1} f_i \sim& \hat{\lambda}_{i,i+1} + \hat{\lambda}_{i,i+m} \\ =& \hat{\lambda}_{i,i+1}(1+\sqrt{\gamma_i}) \\ =& Nu_{i+1}\beta\left(\tfrac{s_{i+1}}{s_i}\right)(1+\sqrt{\gamma_i}) \\ =& u_{i+1}(1+\sqrt{\gamma_i}), \end{aligned} \tag{12.195}$$

where in the last step we used that $s_i = s_{i+1}$ and $\beta(1) = 1/N$. We will extend the argument in the proof of Theorem 3 in Durrett et al. [20], and indicate that the total forward rate of fixation from i should rather be

$$Nu_{i+1} f_i \sim \hat{\lambda}_{i,i+1}\chi\left(\frac{\gamma_i}{\beta(s_m)}\right) = u_{i+1}\chi\left(\frac{\gamma_i}{\beta(s_m)}\right), \tag{12.196}$$

where $\chi(\cdot)$ is the function defined in (12.63). This will also motivate (12.114), since this formula serves the purpose of modifying the incorrect forward rate of fixation (12.195), so that it equals the adjusted one in (12.196), keeping the relative sizes of the different forward rates $i \to j$ of fixation intact for $j = i+1, \ldots, m$.

The rationale for (12.196) is that type $i+1$ mutations arrive according to a Poisson process at rate Nu_{i+1}, and χ/N is the probability that any such type $i+1$ mutation has descendants of type $i+1$ or m that spread to the whole population. We need to show that

$$\chi = \chi\left(\frac{\gamma_i}{\beta(s_m)}\right). \tag{12.197}$$

To this end, let X_t be the fraction of descendants of a $i \to i+1$ mutation, Nt time units after this mutation appeared. We stop this process at a time point τ when X_t reaches any of the two boundary points 0 or 1 ($X_\tau = 0$ or 1), or when a successful mutation $i+1 \to i+2$ appears before that, which is a descendant of the type $i+1$ mutation that itself will have type m descendants who spread to the whole population, before any other type gets fixed ($0 < X_\tau < 1$). We have that $x = X_0 = 1/N$, but define

$$\bar{\beta}(s_m; x) = \bar{\beta}(x) = P(X_\tau = 0 | X_0 = x)$$

for any value of x. This is a non-fixation probability, i.e. the probability that the descendants of Nx individuals of type $i+1$ at time $t = 0$ neither have a successful type $i+2$ descendant, nor take over the population before that. Since the descendants of a single type $i+1$ mutation take over the population with probability $1 - \bar{\beta}(1/N)$, it is clear that

$$\chi = N\left[1 - \bar{\beta}\left(\frac{1}{N}\right)\right] \sim \lim_{x\to 0} \frac{1-\bar{\beta}(x)}{x} = -\bar{\beta}'(0). \tag{12.198}$$

Durrett et al. [20] prove that it is possible to neglect the impact of further $i \to i+1$ mutations after time $t = 0$. It follows that X_t will be a version of the Moran process of Appendix B with $s = s_{i+1}/s_i = 1$, during the time interval $(0, \tau)$, when time speeded up by a factor of N. Using (12.123)–(12.124), we find that the infinitesimal mean and variance functions of X_t are

$$\begin{aligned} M(x) =& N \times 0 = 0, \\ V(x) =& N \times 2x(1-x)/N = 2x(1-x), \end{aligned} \tag{12.199}$$

respectively. At time t, a successful type $i+2$ mutation arrives at rate

$$\begin{aligned} N \times NX_t \times u_{i+2} q_{i+1,m} \beta\left(\tfrac{s_m}{s_i}\right) \sim& N^2 X_t \times u_{i+2} r_{i+1,m} \beta(s_m) \\ =& N^2 X_t \times r_{im}^2 \beta(s_m) \\ =& X_t \times (\hat{\lambda}_{im}/\hat{\lambda}_{i,i+1})^2 \beta(s_m)^{-1} \\ =& X_t \times \gamma_i \beta(s_m)^{-1} \\ =:& X_t \times \gamma', \end{aligned} \tag{12.200}$$

where in the second step we used $r_{im}^2 = u_{i+2} r_{i+1,m}$, which follows from (12.36), since all $R(\rho_{ilj}) = 1$ when $s_1 = \cdots = s_{m-1} = 1$. Then in the third step we used $\hat{\lambda}_{im}/\hat{\lambda}_{i,i+1} = N r_{im} \beta(s_m)$, which follows from (12.35), and in the last step we introduced the short notation $\gamma' = \gamma_i \beta(s_m)^{-1}$. (One instance of γ' is presented for the boundary scenarios of Sect. 12.7.2.1, below formula (12.105).)

We will use (12.199)–(12.200) and Kolmogorov's backward equation in order to derive a differential equation for $\bar{\beta}(x)$. Consider a fixed $0 < x < 1$, and let $h > 0$ be a small number. Then condition on what happens during time interval $(0, h)$. When h is small, it is unlikely that the process X_t will stop because it hits any of the boundaries 0 or 1, i.e.

$$\begin{aligned} P(\tau < h, 0 < X_\tau < 1) =& x\gamma' h + o(h), \\ P(\tau < h, X_\tau \in \{0, 1\}) =& o(h) \end{aligned}$$

as $h \to 0$. The non-fixation probability can therefore be expressed as

$$\begin{aligned} \bar{\beta}(x) =& x\gamma' h \times 0 + (1 - x\gamma' h) \int_0^t \bar{\beta}(y) dP(X_h = y | X_0 = x) + o(h) \\ =& (1 - x\gamma' h) \left[\bar{\beta}(x) + \tfrac{1}{2} V(x) \bar{\beta}''(x) h \right] + o(h). \end{aligned}$$

Letting $h \to 0$, we find from (12.199) that $\bar{\beta}(x)$ satisfies the differential equation

$$x(1 - x)\bar{\beta}''(x) - x\gamma' \bar{\beta}(x) = 0. \tag{12.201}$$

Durrett et al. [20] use a power series argument to prove that the solution of (12.201), with boundary conditions $\bar{\beta}(0) = 1$ and $\bar{\beta}(1) = 0$, is

$$\bar{\beta}(x) = \frac{\sum_{k=1}^{\infty} \frac{(\gamma')^k}{k!(k-1)!} (1 - x)^k}{\sum_{k=1}^{\infty} \frac{(\gamma')^k}{k!(k-1)!}}. \tag{12.202}$$

Recalling (12.63) and that $\gamma' = \gamma_i / \beta(s_m)$, we deduce formula (12.197) from (12.198) and differentiation of (12.202) with respect to x. □

References

1. Asmussen, S., Nerman, O., Olsson, M.: Fitting phase-type distributions via the EM algorithm. Scand. J. Stat. **23**, 419–441 (1996)
2. Axe, D.D.: The limits of complex adaptation: an analysis based on a simple model of structured bacterial populations. BIO-Complex. **2010**(4) (2010)
3. Barton, N.H.: The probability of fixation of a favoured allele in a subdivided population. Genet. Res. **62**, 149–158 (1993)
4. Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K.W., Velculescu, V.W., Vogelstein, B., Nowak, M.A.: Genetic progression and the waiting time to cancer. PLoS Comput. Biol. **3**(11), e225 (2007)

5. Behe, M., Snoke, D.W.: Simulating evolution by gene duplication of protein features that require multiple amino acid residues. Protein Sci. **13**, 2651–2664 (2004)
6. Behe, M., Snoke, D.W.: A response to Michael Lynch. Protein Sci. **14**, 2226–2227 (2005)
7. Behrens, S., Vingron, M.: Studying evolution of promoter sequences: a waiting time problem. J. Comput. Biol. **17**(12), 1591–1606 (2010)
8. Behrens, S., Nicaud, C., Nicodéme, P.: An automaton approach for waiting times in DNA evolution. J. Comput. Biol. **19**(5), 550–562 (2012)
9. Bobbio, A., Horvath, Á., Scarpa, M., Telek, M.: A cyclic discrete phase type distributions: properties and a parameter estimation algorithm. Perform. Eval. **54**, 1–32 (2003)
10. Bodmer, W.F.: The evolutionary significance of recombination in prokaryotes. Symp. Soc. General Microbiol. **20**, 279–294 (1970)
11. Carter, A.J.R., Wagner, G.P.: Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. Proc. R. Soc. Lond. **269**, 953–960 (2002)
12. Cao, Y., et al.: Efficient step size selection for the tau-leaping simulation method. J. Chem. Phys. **124**, 44109–44119 (2006)
13. Chatterjee, K., Pavlogiannis, A., Adlam, B., Nowak, M.A.: The time scale of evolutionary innovation. PLOS Comput. Biol. **10**(9), d1003818 (2014)
14. Christiansen, F.B., Otto, S.P., Bergman, A., Feldman, M.W.: Waiting time with and without recombination: the time to production of a double mutant. Theor. Popul. Biol. **53**, 199–215 (1998)
15. Crow, J.F., Kimura, M.: An Introduction to Population Genetics Theory. The Blackburn Press, Caldwell (1970)
16. Desai, M.M., Fisher, D.S.: Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics **176**, 1759–1798 (2007)
17. Durrett, R.: Probability Models for DNA Sequence Evolution. Springer, New York (2008)
18. Durrett, R., Schmidt, D.: Waiting for regulatory sequences to appear. Ann. Appl. Probab. **17**(1), 1–32 (2007)
19. Durrett, R., Schmidt, D.: Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. Genetics **180**, 1501–1509 (2008)
20. Durrett, R., Schmidt, D., Schweinsberg, J.: A waiting time problem arising from the study of multi-stage carinogenesis. Ann. Appl. Probab. **19**(2), 676–718 (2009)
21. Ewens, W.J.: Mathematical Population Genetics. I. Theoretical Introduction. Springer, New York (2004)
22. Fisher, R.A.: On the dominance ratio. Proc. R. Soc. Edinb. **42**, 321–341 (1922)
23. Fisher, R.A.: The Genetical Theory of Natural Selection. Oxford University Press, Oxford (1930)
24. Gerstung, M., Beerenwinkel, N.: Waiting time models of cancer progression. Math. Popul. Stud. **20**(3), 115–135 (2010)
25. Gillespie, D.T.: Approximate accelerated simulation of chemically reacting systems. J. Chem. Phys. **115**, 1716–1733 (2001)
26. Gillespie, J.H.: Molecular evolution over the mutational landscape. Evolution **38**(5), 1116–1129 (1984)
27. Gillespie, J.H.: The role of population size in molecular evolution. Theor. Popul. Biol. **55**, 145–156 (1999)
28. Greven, A., Pfaffelhuber, C., Pokalyuk, A., Wakolbinger, A.: The fixation time of a strongly beneficial allele in a structured population. Electron. J. Probab. **21**(61), 1–42 (2016)
29. Gut, A.: An Intermediate Course in Probability. Springer, New York (1995)
30. Haldane, J.B.S.: A mathematical theory of natural and artificial selection. Part V: selection and mutation. Math. Proc. Camb. Philos. Soc. **23**, 838–844 (1927)
31. Hössjer, O., Tyvand, P.A., Miloh, T.: Exact Markov chain and approximate diffusion solution for haploid genetic drift with one-way mutation. Math. Biosci. **272**, 100–112 (2016)
32. Iwasa, Y., Michor, F., Nowak, M.: Stochastic tunnels in evolutionary dynamics. Genetics **166**, 1571–1579 (2004)

33. Iwasa, Y., Michor, F., Komarova, N.L., Nowak, M.: Population genetics of tumor suppressor genes. J. Theor. Biol. **233**, 15–23 (2005)
34. Kimura, M.: Some problems of stochastic processes in genetics. Ann. Math. Stat. **28**, 882–901 (1957)
35. Kimura, M.: On the probability of fixation of mutant genes in a population. Genetics **47**, 713–719 (1962)
36. Kimura, M.: Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical and pseudo-sampling methods. Proc. Natl. Acad. Sci. USA **77**, 522–526 (1980)
37. Kimura, M.: The role of compensatory neutral mutations in molecular evolution. J. Genet. **64**(1), 7–19 (1985)
38. Kimura, M., Ohta, T.: The average number of generations until fixation of a mutant gene in a finite population. Genetics **61**, 763–771 (1969)
39. Knudson, A.G.: Two genetic hits (more or less) to cancer. Nat. Rev. Cancer **1**, 157–162 (2001)
40. Komarova, N.L., Sengupta, A., Nowak, M.: Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. J. Theor. Biol. **223**, 433–450 (2003)
41. Lambert, A.: Probability of fixation under weak selection: a branching process unifying approach. Theor. Popul. Biol. **69**(4), 419–441 (2006)
42. Li, T.: Analysis of explicit tau-leaping schemes for simulating chemically reacting systems. Multiscale Model. Simul. **6**, 417–436 (2007)
43. Lynch, M.: Simple evolutionary pathways to complex proteins. Protein Sci. **14**, 2217–2225 (2005)
44. Lynch, M., Abegg, A.: The rate of establishment of complex adaptations. Mol. Biol. Evol. **27**(6), 1404–1414 (2010)
45. MacArthur, S., Brockfield, J.F.Y.: Expected rates and modes of evolution of enhancer sequences. Mol. Biol. Evol. **21**(6), 1064–1073 (2004)
46. Maruyama, T.: On the fixation probability of mutant genes in a subdivided population. Genet. Res. **15**, 221–225 (1970)
47. Maruyama, T., Kimura, M.: Some methods for treating continuous stochastic processes in population genetics. Jpn. J. Genet. **46**(6), 407–410 (1971)
48. Maruyama, T., Kimura, M.: A note on the speed of gene frequency changes in reverse direction in a finite population. Evolution **28**, 161–163 (1974)
49. Moran, P.A.P.: Random processes in genetics. Proc. Camb. Philos. Soc. **54**, 60–71 (1958)
50. Neuts, M.F.: Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. John Hopkins University Press, Baltimore (1981)
51. Nicodéme, P.: Revisiting waiting times in DNA evolution (2012). arXiv:1205.6420v1
52. Nowak, M.A.: Evolutionary Dynamics: Exploring the Equations of Life. Belknap Press, Cambridge (2006)
53. Phillips, P.C.: Waiting for a compensatory mutation: phase zero of the shifting balance process. Genet. Res. **67**, 271–283 (1996)
54. Radmacher, M.D., Kelsoe, G., Kepler, T.B.: Predicted and inferred waiting times for key mutations in the germinal centre reaction: evidence for stochasticity in selection. Immunol. Cell Biol. **76**, 373–381 (1998)
55. Rupe, C.L., Sanford, J.C.: Using simulation to better understand fixation rates, and establishment of a new principle: Haldane's Ratchet. In: Horstmeyer, M. (ed.) Proceedings of the Seventh International Conference of Creationism. Creation Science Fellowship, Pittsburgh, PA (2013)
56. Sanford, J., Baumgardner, J., Brewer, W., Gibson, P., Remine, W.: Mendel's accountant: a biologically realistic forward-time population genetics program. Scalable Comput.: Pract. Exp. **8**(2), 147–165 (2007)
57. Sanford, J., Brewer, W., Smith, F., Baumgardner, J.: The waiting time problem in a model hominin population. Theor. Biol. Med. Model. **12**, 18 (2015)
58. Schinazi, R.B.: A stochastic model of cancer risk. Genetics **174**, 545–547 (2006)

59. Schinazi, R.B.: The waiting time for a second mutation: an alternative to the Moran model. Phys. A. Stat. Mech. Appl. **401**, 224–227 (2014)
60. Schweinsberg, J.: The waiting time for *m* mutations. Electron. J. Probab. **13**(52), 1442–1478 (2008)
61. Slatkin, M.: Fixation probabilities and fixation times in a subdivided population. Evolution **35**, 477–488 (1981)
62. Stephan, W.: The rate of compensatory evolution. Genetics **144**, 419–426 (1996)
63. Stone, J.R., Wray, G.A.: Rapid evolution of *cis*-regulatory sequences via local point mutations. Mol. Biol. Evol. **18**, 1764–1770 (2001)
64. Tuğrul, M., Paixão, T., Barton, N.H., Tkačik, G.: Dynamics of transcription factor analysis. PLOS Genet. **11**(11), e1005639 (2015)
65. Whitlock, M.C.: Fixation probability and time in subdivided populations. Genetics **164**, 767–779 (2003)
66. Wodarz, D., Komarova, N.L.: Computational Biology of Cancer. Lecture Notes and Mathematical Modeling. World Scientific, New Jersey (2005)
67. Wright, S.: Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)
68. Wright, S.: The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings of the 6th International Congress on Genetics, vol. 1, pp. 356–366 (1932)
69. Wright, S.: Statistical genetics and evolution. Bull. Am. Math. Soc. **48**, 223–246 (1942)
70. Yona, A.H., Alm, E.J., Gore, J.: Random sequences rapidly evolve into *de novo* promoters (2017). bioRxiv.org, https://doi.org/10.1101/111880
71. Zhu, T., Hu, Y., Ma, Z.-M., Zhang, D.-X., Li, T.: Efficient simulation under a population genetics model of carcinogenesis. Bioinformatics **6**(27), 837–843 (2011)

Chapter 13
Characterizing the Initial Phase of Epidemic Growth on Some Empirical Networks

Kristoffer Spricer and Pieter Trapman

Abstract A key parameter in models for the spread of infectious diseases is the basic reproduction number R_0, which is the expected number of secondary cases a typical infected primary case infects during its infectious period in a large mostly susceptible population. In order for this quantity to be meaningful, the initial expected growth of the number of infectious individuals in the large-population limit should be exponential. We investigate to what extent this assumption is valid by simulating epidemics on empirical networks and by fitting the initial phase of each epidemic to a generalised growth model, allowing for estimating the shape of the growth. For reference, this is repeated on some elementary graphs, for which the early epidemic behaviour is known. We find that for the empirical networks tested in this paper, exponential growth characterizes the early stages of the epidemic, except when the network is restricted by a strong low-dimensional spacial constraint.

Keywords Epidemics · Exponential growth · Generalized growth model
Reproduction number · Stochastic processes

13.1 Introduction

A key parameter in many mathematical models that describe the spread of infectious diseases is the basic reproduction number R_0. It may be understood as the expected number of other individuals a typical infected individual infects during his/her infectious period in a large mostly susceptible population (see [4, p. 4]). The basic reproduction number serves as a threshold parameter, in the sense that in most standard models, if $R_0 \leq 1$ a large outbreak is impossible, while if $R_0 > 1$, a large outbreak occurs with positive probability. In those models, preventing a frac-

K. Spricer (✉) · P. Trapman
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
e-mail: spricer@math.su.se

P. Trapman
e-mail: ptrapman@math.su.se

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_13

tion $1 - 1/R_0$ of the infections (e.g. through vaccination) is enough to stop a major outbreak (see [4, p. 209]).

The above properties of R_0 are strongly connected to the correspondence of R_0 with the offspring mean of a branching process approximation of the epidemic. So for R_0 to be meaningful, the initial expected growth of the number of infectious individuals in the large-population limit should be exponential. This exponential growth is present in SIR epidemics (Susceptible → Infectious → Recovered; a definition is given in Sect. 13.2.3) in large homogeneously mixing populations, in which all individuals have the same characteristics and all pairs of individuals independently make contacts with the same rate. It is also present in many well-studied generalizations of this SIR model in large homogeneously mixing populations. Generalizations are e.g. possible through SIS (Susceptible → Infectious → Susceptible) or SIRS (Susceptible → Infectious → Recovered → Susceptible) models, or models with demographic turnover through births, deaths and migration. Other generalizations are allowing for heterogeneity among the individuals and contact rates between pairs, e.g. through allowing for household structures, multi-type structures and some network structures in the population (see e.g. [4] for descriptions of these models and population structures). Even with these generalizations, major outbreaks of epidemics still show exponential growth in the initial phase of the epidemic and therefore R_0 is a meaningful parameter (see [23] and references therein).

A trade-off between realism and analytical tractability is often necessary in developing a mathematical model. Because of the reasons stated above, in many instances this tractability requires the possibility of exponential growth in the model, either directly or as a byproduct of other assumptions. It is not a-priori clear in which cases real-life spread of infectious diseases allows for a meaningful definition of R_0 and in which cases the use of R_0 may be misleading and other key parameters should be estimated. For example, it is well known that SIR-epidemics on essentially 2-dimensional networks grow linearly whenever contacts between vertices are mostly local, i.e. if the probability of long range contacts decays sufficiently fast. The epidemic then spreads in the form of travelling waves on the plane (see e.g. [15, 25], but also [7, 22] for models where long-range contacts change the behaviour of the spread). Human physical activity is mostly restricted to the 2-dimensional nature of the earth's surface and a natural assumption is that graphs based on human interactions (e.g. in social networks) may also show this restriction. Although, in this paper, we focus on SIR epidemics, we expect that our discussion also applies to rumours, evolution or games on the networks. Furthermore, quantities of interest such as the diameter and typical distances in networks (see [10, Chapter 1]) are related to the possibility of exponential growth of an epidemic on the network, and in this context the typical distances are also strongly related to the so-called "six degrees of separation" and "small-world" phenomena (see [26]).

In the present paper we study (simulated) SIR epidemics on several theoretical and empirical networks and investigate to what extent they exhibit exponential and subexponential growth. All, but one, of the empirical networks are taken from [13] and we are aware that those networks are at best a proxy for networks relevant for the spread of infectious diseases. Throughout the analysis, our key assumption is that

if the number of vertices in such random graphs goes to infinity and if exponential growth is possible, then at some level (see [2]) a branching process approximation is possible and R_0 is well defined. This R_0 should then be estimated from the initial phase of a (simulated) epidemic on the finite empirical network. Thus, we analyze the development of the epidemic during the first generations (with exception of the first generation) by fitting the development of the epidemic to a *generalized growth model* (see Sect. 13.2.5). The process is repeated for many simulated epidemics on each graph and the collective information from the analysis of these epidemics is used to compare the initial growth of epidemics on different graphs (see Sect. 13.3).

In most of the analyzed empirical networks the unrestricted epidemic, where an infected vertex infects *all* susceptible neighbours, grows so fast that the initial phase of the epidemic is over in as few as 3–4 generations. To be able to study the epidemic for more generations, we restrict it by two different methods described in Sect. 13.2.4. These restrictions reduce the reproduction number so that it takes longer before a substantial portion of the population has been infected. Since such restrictions could affect the way in which the epidemic spreads in addition to just slowing it down, a secondary objective is to study such effects. A discussion can be found in Sect. 13.4.

13.2 Model

In this section we present the models for epidemic growth on graphs that we explore in the paper, expanding on some of the concepts introduced in Sect. 13.1. We start with definitions and results concerning graphs in Sect. 13.2.1. An overview of the specific graphs that we analyze in this paper is given in Sect. 13.2.2. In Sect. 13.2.3 we give a brief account of the SIR-model on graphs in *discrete time* (i.e. in a generation perspective) and discuss what we mean by the growth of the epidemic, specifically by exponential growth. We describe the two methods we use to "slow the epidemic down" in Sect. 13.2.4. The method used to analyze the growth of the simulated epidemics is presented in Sect. 13.2.5.

13.2.1 Graphs

A finite graph is a set of n vertices together with a set of edges that join vertices pairwise. We consider *simple* and *undirected* graphs, i.e. there are no self-loops (an edge connecting a vertex to itself) and no parallel edges (several edges join the same pair of vertices) and all edges are undirected (see [10]). When considering epidemics on graphs, we let *vertices* represent people and *edges* represent relationships between people through which the infectious disease may spread, and the infection can spread in both directions via the undirected edges. For any given vertex the vertices that can be reached directly through an edge are called the *neighbours* of the vertex. The *degree* of vertex v is the number of neighbours of v. We denote this degree by d_v. For

a given graph we talk about the degree sequence $\mathbf{d} = (d_1, d_2, \ldots, d_n)$. Without loss of generality we restrict the analysis to graphs where $d_i \geq 1$ for all $i = 1, 2, \ldots, n$, since vertices with degree 0 cannot interact in an epidemic and are therefore not really part of the network. Let Z be the degree of a vertex selected uniformly at random from the graph and define $p_k = \Pr(Z{=}k)$. The distribution of Z is the *degree distribution* of the graph and we define $\mu = \mathrm{E}[Z]$ and $\sigma^2 = \mathrm{Var}(Z)$.

Some graphs are *deterministic*, e.g. the graph on the Euclidean integer lattice $\mathbb{Z}^\eta$, where η is the dimension of the lattice. Here the integer points in $\mathbb{Z}^\eta$ are the vertices and edges exist between all pairs of vertices with Euclidean distance 1. We consider finite subsets (tori) of the infinite graph $\mathbb{Z}^\eta$ in order to make comparisons with other finite graphs. Other graphs are *random* in the sense that they are constructed probabilistically, e.g. the Erdős–Rényi graph (see [5, 6]) and configuration model graphs (see [5, 16]). In both cases the number of vertices is given and finite. For an extensive overview of results on both Erdős–Rényi and configuration model graphs see [5, 10].

In the Erdős–Rényi graph there exists an edge between any two vertices with probability $\frac{\lambda}{n-1}$, where λ is a given constant equal to the expected degree of a vertex selected uniformly at random, and the presence or absence of possible edges are independent. This construction results in the degree distribution of a vertex selected uniformly at random being the binomial distribution with parameters $n-1$ and $\frac{\lambda}{n-1}$. For $n \to \infty$, the degree distribution converges to a Poisson distribution with parameter λ.

In the configuration model graph, we either start with a given degree sequence $\mathbf{d}$ (which may be taken from an empirical network) for the vertices, or the degree sequence is independent and identically distributed (i.i.d.), with given distribution D (which may also be taken from an empirical network). We then create the graph as follows: To each (for the moment unconnected) vertex we assign a number of "stubs" corresponding to its degree. The stubs are paired uniformly at random to create edges. A possible left over stub is deleted and so are any self-loops, while parallel edges are merged to one edge. So, the created graph is simple. If the degree sequence is an i.i.d. sequence with distribution D, and if D has finite mean, then in the limit $n \to \infty$ the degree distribution of the obtained configuration model graph converges in probability to the distribution D (see [3]).

Empirical graphs are created from real world data, e.g. from observed social interactions within a group of people. These networks have a given finite size, while we are interested in asymptotic results when the population size $n \to \infty$. Therefore, as pointed out in Sect. 13.1, we analyze (processes on) the empirical graph as if the empirical graph is a realization of a random graph, which can be analyzed for $n \to \infty$. The exact mechanism of constructing the random graph is typically unknown. The empirical graphs we have analyzed in this paper are described in Sect. 13.2.2, below.

In our analysis, we treat the empirical networks as if they are realizations of some unspecified random graph model, which can be defined for an arbitrary large number of nodes in the network. By studying epidemics on these realizations we try to answer whether, in the large population limit of the random graph, exponential growth is possible. However, we only have access to a limited number of such realizations

and we cannot freely control the number of nodes in them. Also the real dynamics through which an empirical network is created are probably too hard to describe and analyze and possibly not even random. Therefore mathematical models and results based on such empirical realizations should be interpreted with care.

13.2.2 The Studied Networks

In this subsection we present the networks we use to generate the graphs that are analyzed in this paper and we discuss some of their properties.

A summary of properties of the networks used in this paper can be found in Table 13.1. The first three networks in the table are from the Stanford Large Dataset Collection (see [13]).

The graphs are discussed below.

- **soc-LiveJournal1** is a large online social network that allows for the formation of communities. On the network people state who their "friends" are and although this does not have to be mutual, it often is. In our model we only consider the mutual statements of friendship and let these be represented by undirected edges, while people are represented by vertices. In figures the graph is referred to as "LJ".
- **ca-CondMat** is based on the *arXiv* condensed matter collaboration network (COND-MAT). Authors are represented as vertices and undirected edges are present between all authors that are listed as co-authors of the same paper. In figures the graph is referred to as "CM".
- **roadNet-PA** is based on the road network of Pennsylvania. Intersections between roads are represented by vertices and roads are represented by edges. Because of the spatial nature of a road network we expect to see spatial restrictions in this network and this is why it was included in the analysis. In figures the graph is referred to as "Rd".

Table 13.1 An overview of networks that are investigated in this paper. The number of undirected edges and the number of vertices with at least one edge are indicated

Data set	# vertices	# edges	Type of graph
soc-LiveJournal1	3 823 816	25 624 154	Online social network
ca-CondMat	23 133	93 439	Scientific collaboration network
roadNet-PA	1 088 092	1 541 898	Road network in Pennsylvania
Swedish population	7 616 569	18 139 894	Workplace and family
D2	1 000 000	2 000 000	2-dimensional lattice
D6	1 000 000	6 000 000	6-dimensional lattice

- **Swedish population**[1] is a large network that is based on data containing only the workplace and family affiliation of people in Sweden (see also [11]). Although the used dataset does not contain geographic information, it may still be assumed that family location and workplace location can be spatially correlated, thus imposing a spatial structure on the entire graph.
  Because some of the workplaces are large, we have assumed that people interact with colleagues only in smaller working groups. We model this by (randomly) dividing the workplaces into groups of 7 people (with at most one group in each company having a size between 1 and 6 when the company size is not divisible by 7).
  A *reference* version of this dataset was also tested where company affiliation was assigned at random to each vertex, while keeping the distribution of workplace sizes fixed. If epidemics on this reference graph differ from the original graph, this could be an indication of spatial restrictions on the original graph. In figures the graphs are referred to as "Sw" and "SR", respectively.
- **D2** is a finite regular square lattice (on $\mathbb{Z}^2$) in the shape of a torus with sides of 10^3 vertices, thus in total 10^6 vertices. Because of the torus shape there is no center in the graph and the development of the epidemic does thus not depend on where the epidemic starts. In figures the graph is referred to as "D2".
- **D6** is a finite regular lattice on $\mathbb{Z}^6$ in the shape of a torus with sides of 10 vertices, thus in total 10^6 vertices. As for the D2-graph there is no center in the graph. In figures the graph is referred to as "D6".

13.2.3 Epidemics on Graphs

In this paper we consider SIR-epidemics, where each vertex is either *susceptible*, *infected* (and infectious) or *recovered*. A vertex that is recovered is immune and can never be infected again. We restrict the analysis to epidemics in *discrete time* and also assume that each infected vertex stays infected for only one time unit before it recovers and cannot spread the infection further. This model corresponds to the so-called Reed–Frost model on graphs (see [4, p. 48]). The above implies that, for any finite graph, the epidemic must eventually end when there are no more infected vertices left. The total number of vertices that have been infected during the course of the epidemic is called *the final size* of the epidemic.

The first vertex to be infected is called the *index case*. We assume that the index case was infected at time $i = 0$, where time represents the generation number. The index case then spreads the infection to (a subset of) its neighbours and they in turn spread it to (a subset of) their neighbours. For each generation i, we keep track of I_i, the number of infected vertices in generation i, and of $J_i = \sum_{k=0}^{i} I_k$, the total number of infected vertices up to and including generation i. Note that, *if* the infection always

[1] Data kindly supplied by Fredrik Liljeros, Department of Sociology, Stockholm University.

spreads to *all* neighbours, J_i is equal to the number of vertices within graph distance i from the index case.

A measure of the rate at which the epidemic is growing is the *instantaneous reproduction number* at time (generation) i, which we define as the *average* number of offspring of a vertex in generation $i-1$

$$\overline{m}_i = \frac{I_i}{I_{i-1}}, \tag{13.1}$$

for $i \geq 1$ and conditioned on $I_{i-1} > 0$. The instantaneous reproduction number depends on how many neighbours an infected vertex has, on how many of the neighbours that are still susceptible and on the mechanism by which the vertex infects its neighbours. In the initial phase of an epidemic $\overline{m}_i$ may be approximately constant as a function of i, but for a finite graph it must eventually decrease as there are fewer and fewer susceptible vertices left. If vertices have different degrees or if vertices have different local environments, then the development of the epidemic also depends on which vertex is the index case.

On the square lattice, $\mathbb{Z}^2$, the growth of an unrestricted epidemic (i.e. when an infected vertex infects *all* susceptible friends) is initially linear with $I_i = 4i$ and $\overline{m}_i = \frac{i}{i-1}$, $i \geq 1$ (conditioned on $I_0 = 1$).

On a configuration model graph the initial phase of the epidemic (if the first generation is ignored) is well approximated by a *Galton–Watson* branching process in discrete time, where all individuals reproduce independently with offspring distribution X, with $m = \mathrm{E}[X]$. For a Galton–Watson process

$$I_i = \sum_{j=1}^{I_{i-1}} X_{i,j}, \tag{13.2}$$

where $X_{i,j}$ are all independent and distributed as X. We see that

$$\mathrm{E}\left[I_i \mid I_{i-1}\right] = I_{i-1}\mathrm{E}[X] \tag{13.3}$$

and that

$$\mathrm{Var}\left[I_i \mid I_{i-1}\right] = I_{i-1}\mathrm{Var}[X] \tag{13.4}$$

so that both the (conditional) expectation and the (conditional) variance of I_i are proportional to the size of the previous generation. We use these relationships in the analysis of the data, see Sect. 13.2.5.

In the large population limit the expected reproduction number

$$\mathrm{E}[\overline{m}_i \mid I_{i-1} > 0] = m \tag{13.5}$$

does not depend on the epidemic generation. The expected size of the ith generation of the epidemic is $\mathrm{E}[I_i] = m^i$ (given a single index case and assuming the branching

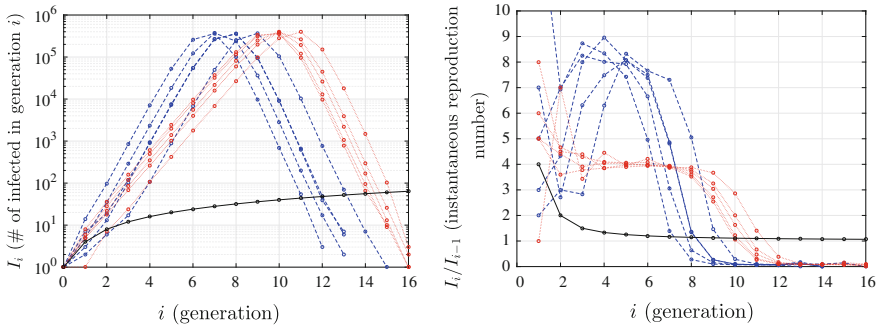

Fig. 13.1 The number of infected vertices (left) and the instantaneous reproduction number (right) as a function of the epidemic generation for three network models—the square lattice (solid black line), the Erdős–Rényi model (dotted red line) and the configuration model with a geometric degree distribution with weight on 0 (dashed blue line)—all with mean degree 4. For each plot the size of the population is 10^6 vertices. In each generation infected vertices infect all of their susceptible neighbours. With the chosen axes scaling in the left plot, the data points fall on a straight line if the growth is exponential, corresponding to an approximately constant instantaneous reproduction number in the right plot

process is valid from the first generation) and we see that the growth of the epidemic is exponential if $m > 1$.

As shown in [9, p. 36]

$$\hat{m}_i = \frac{J_i - 1}{J_{i-1}}, \tag{13.6}$$

(assuming that $I_0 = 1$) is a better estimator for m than $\overline{m}_i$. If the branching process approximation is valid first from generation 2 then we can modify Eq. (13.6) slightly to obtain

$$\hat{m}_i = \frac{J_i - J_1}{J_{i-1} - J_0}. \tag{13.7}$$

This latter expression is most relevant in this paper, since we select the index case uniformly at random among all vertices, while subsequent vertices are infected by following edges from an infected vertex. This causes the degree distribution of the index case to differ from that of vertices that are infected later, as we explain now (see also [17]).

The development of epidemics on graphs depends both on local (such as the degree of a vertex) and global structural properties. As an example of the relevance of those structural properties, in Fig. 13.1 we have plotted a few SIR-epidemics that were simulated on three different network types that all have the same mean degree 4. These networks are described in more detail in Sect. 13.2.1. Here we have assumed that infected vertices infect all of their susceptible neighbours. We observe, that although the three graphs have the same mean degree, epidemics on the three graphs develop differently. For the *Erdős–Rényi* model and the *configuration model* early in the epidemic the growth is approximately exponential, while the *square*

lattice exhibits essentially linear growth. In the right figure for the first two graphs this is illustrated by approximately constant *instantaneous reproduction numbers* (the average number of new infections caused per infected vertex in the current infection generation), well above 1 in the initial phase of the epidemic, after which the instantaneous reproduction number drops to a value below 1. For the square lattice model the instantaneous reproduction number drops rapidly from the very start of the epidemic, asymptotically approaching 1. In the latter case the subexponential development is an effect of the spatial structure of the network.

Before continuing, we remind the reader that $\mu = E[Z]$ and $\sigma^2 = \mathrm{Var}(Z)$, where Z is the degree of a vertex that is chosen uniformly at random among all vertices in the graph (see also Sect. 13.2.1). Let $\widetilde{Z}$ be the degree of a vertex that is selected by first selecting an edge uniformly at random and then selecting one of the two connected vertices at random. On the configuration model graph, again ignoring the first generation, initially and for as long as the branching process approximation is valid, the epidemic growth is governed by $X \sim \widetilde{Z}-1$, where

$$\widetilde{p}_k = \Pr\left(\widetilde{Z}{=}k\right) = \frac{kp_k}{\mu}. \tag{13.8}$$

The "-1" is because the infection cannot spread back to "the infector" since it is by definition not susceptible any more. The *expected* reproduction number in the initial phase of the epidemic is thus $m = \widetilde{\mu}-1$, where

$$\widetilde{\mu} = \mathrm{E}\left[\widetilde{Z}\right] = \frac{\mathrm{E}\left[Z^2\right]}{\mathrm{E}[Z]} = \mu + \frac{\sigma^2}{\mu}. \tag{13.9}$$

Thus $\widetilde{\mu}$ can be much larger than μ if σ^2 is much larger than μ.

13.2.4 Restricting the Reproduction Number

Similar to the configuration model, some of the empirical graphs analyzed in this paper have instantaneous reproduction numbers, in the initial phase of the epidemic, that are much larger than the mean degree of the graph. An unrestricted epidemic on such a graph grows very fast, infecting most of the population in just a few generations. This makes it difficult to assess if the growth is exponential or not. To work around this problem, we restrict the epidemic so that it develops slower, giving more generations to analyze. This is reasonable, since in real world epidemics we do not expect that each infected vertex infects all of its neighbours.

We use two methods to restrict the instantaneous reproduction number:

1. *Maximum bound, with replacement*: Every infected vertex distributes c infection attempts uniformly at random *with* replacement among all neighbours (including the one who infected him). Here c is a constant. Thus an infected vertex can infect

0 (if all attempts are with non-susceptible neighbours) up to c of its neighbours (if all attempts are with susceptible ones). If c is sufficiently large (often $c = 2$ is enough), this method (typically) allows for large epidemics to develop since both infected vertices with few and infected vertices with many neighbours have a good chance of infecting other vertices. The method is similar to, but not identical with the method used in [14].
Note that this method creates an asymmetry between vertices, in effect turning the undirected graph into a directed graph: it may be that if vertices v_1 and v_2 are neighbours then it is more likely that v_1 infects v_2 (should v_1 become infected before v_2), than that v_2 infects v_1 (should v_2 become infected first).

2. *Bernoulli thinning*: Each susceptible neighbour is infected with probability p. This method is equivalent to the discrete-time version of the Reed–Frost model, where it is assumed that each infected vertex infects each neighbour with probability p (see [4, p. 48]). This is closely related to bond percolation on the graph (see [8]).
A disadvantage of Bernoulli thinning is that, for the datasets that we analyze, in order to significantly slow down the epidemic p has to be so low that vertices with few neighbours have a high probability of not infecting any other vertex. The epidemic is spread mainly through high degree vertices, resulting in fewer infected vertices and a smaller final size of the epidemic. Eventually this has a negative effect on the number of generations that can be used to estimate the instantaneous reproduction number, counteracting the intention of the Bernoulli thinning.

When one of the above mentioned restrictions is applied, an infected vertex typically infects only a subset of its neighbours and the epidemic develops differently on each realization, even if it starts with the same index case. This introduces randomness even for epidemics on non-random graphs.

In this paper we have chosen to slow the epidemic down in such a way that we obtain a sufficient number of generations to analyze, while still leaving the possibility of having a large epidemic. For this purpose the maximum bound restriction $c = 3$ worked on all graphs. We used this throughout the analysis, unless explicitly stated otherwise. On each graph, for the Bernoulli thinning we then select a value p such that the average reproduction number, over many simulations, early on in the epidemic is close to that of epidemics restricted using maximum bound.

We also restrict the epidemics on the reference graphs. For the configuration model graphs we obtain exact expected reproduction numbers that are valid for as long as the branching process approximation holds. For calculating these expected reproduction numbers using the two restriction methods, we start with the offspring distribution X and derive the expectation of the *restricted* distribution L.

We remind the reader that $\widetilde{Z}$ is the degree distribution of vertices reached early on in an epidemic on a configuration model graph, excluding the index case itself (see Sect. 13.2.3). For Bernoulli thinning, remembering that $X \sim \widetilde{Z}-1$ in the configuration model, we then have that

$$L \mid \widetilde{Z} = k \sim \text{Bin}(k-1, p)$$

so that

$$\text{E}\left[L \mid \widetilde{Z} = k\right] = (k-1)p$$

and

$$\text{E}\left[L \mid \widetilde{Z}\right] = \left(\widetilde{Z} - 1\right) p.$$

Thus

$$\text{E}\,[L] = \text{E}\left[\text{E}\left[L \mid \widetilde{Z}\right]\right] = \text{E}\left[\left(\widetilde{Z} - 1\right) p\right] = (\widetilde{\mu} - 1)\, p = \left(\mu + \frac{\sigma^2}{\mu} - 1\right) p.$$

For maximum bound consider the event $A_i =$ "edge i carries the infection on" (with $i = 1, 2, \ldots, k-1$) and define the indicator variable

$$1_{\{A_i\}} = \begin{cases} 1 & \text{if } A_i, \\ 0 & \text{otherwise.} \end{cases}$$

Then the number of offspring L conditioned on $\widetilde{Z} = k$ becomes

$$L \mid \widetilde{Z}{=}k \;\; = \;\; \sum_{i=1}^{k-1} 1_{\{A_i\}}$$

There are c attempts at carrying the infection on and for each one a neighbour is selected uniformly at random, with replacement. Thus $\Pr(A_i) = 1 - \Pr({A_i}^{\complement}) = 1 - \left(\frac{k-1}{k}\right)^c$, since for each of c attempts the probability is $\frac{k-1}{k}$ that vertex i is not selected to carry the infection on (remembering to deduct the edge on which the infection arrived). Here ${A_i}^{\complement}$ indicates the complement of A_i. Thus

$$\begin{aligned} \text{E}[L \mid \widetilde{Z} = k] &= \text{E}\left[\sum_{i=1}^{k-1} 1_{\{A_i\}}\right] = \sum_{i=1}^{k-1} \text{E}\left[1_{\{A_i\}}\right] = \sum_{i=1}^{k-1} \Pr(A_i) \\ &= \sum_{i=1}^{k-1} \left(1 - \left(\frac{k-1}{k}\right)^c\right) = (k-1)\left(1 - \left(1 - \frac{1}{k}\right)^c\right) \end{aligned}$$

and so

$$\text{E}\left[L \mid \widetilde{Z}\right] = \left(\widetilde{Z} - 1\right)\left(1 - \left(1 - \frac{1}{\widetilde{Z}}\right)^c\right),$$

finally

$$\text{E}[L] = \text{E}\left[\text{E}\left[L \mid \widetilde{Z}\right]\right] = \text{E}\left[\left(\widetilde{Z} - 1\right)\left(1 - \left(1 - \frac{1}{\widetilde{Z}}\right)^c\right)\right]. \tag{13.10}$$

This expression can be simplified for the specific values of c that we focus on in this paper:

$$\mathrm{E}[L] = \begin{cases} 2 - 3\mathrm{E}\left[\frac{1}{\widetilde{Z}}\right] + \mathrm{E}\left[\frac{1}{\widetilde{Z}^2}\right] & \text{when } c = 2, \\ 3 - 6\mathrm{E}\left[\frac{1}{\widetilde{Z}}\right] + 4\mathrm{E}\left[\frac{1}{\widetilde{Z}^2}\right] - \mathrm{E}\left[\frac{1}{\widetilde{Z}^3}\right] & \text{when } c = 3. \end{cases} \tag{13.11}$$

Using

$$\mathrm{E}\left[\frac{1}{\widetilde{Z}^n}\right] = \frac{1}{\mu}\mathrm{E}\left[\frac{1}{Z^{n-1}}\right]$$

this can also be expressed as

$$\mathrm{E}[L] = \begin{cases} 2 - \frac{3}{\mu} + \frac{1}{\mu}\mathrm{E}\left[\frac{1}{Z}\right] & \text{when } c = 2, \\ 3 - \frac{6}{\mu} + \frac{4}{\mu}\mathrm{E}\left[\frac{1}{Z}\right] - \frac{1}{\mu}\mathrm{E}\left[\frac{1}{Z^2}\right] & \text{when } c = 3. \end{cases} \tag{13.12}$$

The branching process approximation of an epidemic on a configuration model graph together with the expected reproduction number of the restricted epidemic can be used as a reference for the empirical graphs.

13.2.5 Estimating the Shape of Growth

Much of the theory on epidemics on graphs is about obtaining results for epidemics on infinite graphs, such as Euclidean lattices (see [8]), or obtaining asymptotic results for a sequence of related epidemics on finite graphs, when the graph size grows to infinity (see e.g. [1]). For example, for a configuration model graph (discussed more in Sect. 13.2.1) with n vertices, the initial growth of an epidemic can be analyzed using a branching process approximation (see [1, 2, 12]), which can be shown to be exact (under some extra conditions) until roughly $\sqrt{n}$ vertices have been infected, with probability tending to 1 if $n \to \infty$. This follows from a *birthday problem* type argument (see e.g. [4, p. 54]).

For branching processes it is known that, if the branching process survives, the growth is almost surely asymptotically exponential (see [12]), and the instantaneous reproduction number converges to its expectation, which is the basic reproduction number R_0. If a substantial fraction of the vertices have been infected in an SIR epidemic on a finite graph, with high probability, some neighbours of newly infected vertices have already been infected before, thus reducing the instantaneous reproduction number. Exponential growth, if it exists, is thus only visible in the initial phase of the epidemic, when only a small portion of the graph has been infected. We thus restrict our analysis to the initial phase of the epidemic.

A direct method of analyzing the epidemic growth is to look at how I_i develops over the epidemic generations i. From the discussion of the expected growth shape

for some graph types in Sect. 13.2.3 we conclude that we need a method that is able to handle shapes from linear to exponential. One way to do this is to use a function that models the highest polynomial degree of the growth curve

$$I_i = \alpha(i + \beta)^\gamma, \tag{13.13}$$

where α, β and γ are parameters that we determine by fitting the function to our data. In this paper γ is the parameter of interest. We expect it to be close to 1 if the growth is linear (as for the square lattice) and it should be substantially higher than 1 if the growth is exponential (as for the configuration model). The parameter β is introduced since we do not expect the first generation to show the same expected growth as subsequent generations (as discussed in Sect. 13.2.3). We thus ignore the first generation and allow for some offset for the time i. Unfortunately, because the parameters are highly dependent, the chosen parametrization in Eq. (13.13) does not give good convergence when using standard methods of fitting the equation to data.

An alternative parametrization was originally suggested in [21] in the context of *superexponential* growth and was used to study the impact of superexponential population growth on genetic variations in [18]. The same method was used for *subexponential* growth in [24]. The parametrization was developed for continuous time applications, but can be adapted to our discrete time data.

The basic idea in [21] is to start with a differential equation with two parameters

$$\frac{\mathrm{d}f(t)}{\mathrm{d}t} = rf(t)^a, \tag{13.14}$$

where $f(t)$ can be viewed as modelling the total population size or the number of infected (depending on application) at time t and r and a are parameters. $f(t)$ in continuous time corresponds to I_i in discrete time simply by setting $I_i = f(i)$. a defines the *shape* of the growth curve, while r is a proportionality constant which we may interpret as a measure of the *rate* of growth. The solution to Eq. (13.14) depends on the value of a:

$$f(t) = \begin{cases} be^{rt} & \text{if a} = 1 \quad (13.15a) \\ \left(r(1-a)t + b^{1-a}\right)^{\frac{1}{1-a}} & \text{otherwise,} \quad (13.15b) \end{cases}$$

where $b = f(0)$ is given by the starting condition, the size of the population at time 0. We note that Eq. (13.15a) is the limit of Eq. (13.15b) as $a \to 1$. When $a = 1$ we recognize that r is the *Malthusian parameter* (see e.g. [4, p. 10]). We also note that Eq. (13.15b) is essentially a reparametrization of Eq. (13.13).

Taking $I_i = f(i)$ in Eq. (13.15b) we obtain

$$I_i = \left(r(1-a)i + b^{1-a}\right)^{\frac{1}{1-a}}. \tag{13.16}$$

When performing the fit we take into account that the variance of I_i is not constant. Rather, we can expect it to increase if the generation size is larger and in our model we go further and assume that it is approximately *proportional* to the generation size. This is reasonable considering that the conditional variance and the conditional expectation of the generation size are proportional to each other (as noted in Sect. 13.2.3). This lends itself well to a log-transformation to obtain

$$\log(I_i) = \frac{1}{1-a} \log\left(r(1-a)i + b^{1-a}\right). \tag{13.17}$$

Although this model has a singular point when the growth is exactly exponential ($a = 1$), this case is unlikely with empirical data and we choose to ignore the singular point and use Eq. (13.16) as it is.

In the model, I_i is the data that is obtained from each individual simulated epidemic and a, r and b are treated as unknown parameters. By fitting Eq. (13.17) to the data we obtain estimates of the parameter triple (a, r, b). The fit is performed using least squares regression by supplying Eq. (13.17) as a custom function to the *fit*-function in *Matlab* (see [20]). The *fit*-function is supplied with starting points (0.5, 0.5, 0.5), minimum allowed values $(-10, 0, 0)$ and maximum allowed values $(5, 10^5, 100)$ for the parameter triple. In addition, R^2_{adj} (the adjusted coefficient of determination, see e.g. [19, p. 433]), produced by the *fit*-function, was inspected, but the value was not used to discard any results. R^2_{adj} were in general high, except for epidemics on the road network. These depart most from the shape assumed by the generalized growth model and this is also reflected in the large variation in parameter estimates that can be seen in Fig. 13.2.

Conditioned on having a *good fit*, we can then interpret a as a measure of how linear or how exponential the epidemic growth is. Values of a close to 1 can be interpreted as having exponential growth, while values close to 0 correspond to linear growth, such as we expect for the square lattice. Negative values correspond to sub-linear growth.

How good the fit needs to be to draw conclusions about a single epidemic depends on the application. However, through simulation we have access to many epidemics from each graph. Thus we can assess how similar or how different graphs are by comparing the parameter estimates from a large number of simulated epidemics for each graph. As already stated in [24] this is a phenomenological approach and as such it does not properly justify why this specific model and parametrization of the growth curve should be used. We justify the method by also simulating epidemics on known (reference) graphs and by using parameter estimates from those. Our reference graphs are regular lattices and configuration model graphs. We interpret the parameter estimates from the empirical graphs with respect to those obtained on the reference graphs.

The branching process approximation discussed in Sect. 13.2.3 works well until there is a substantial probability that an infected vertex tries to infect an already immune vertex. Given that J is the total number of vertices that has been infected in the epidemic, this probability would in a configuration model be approximately

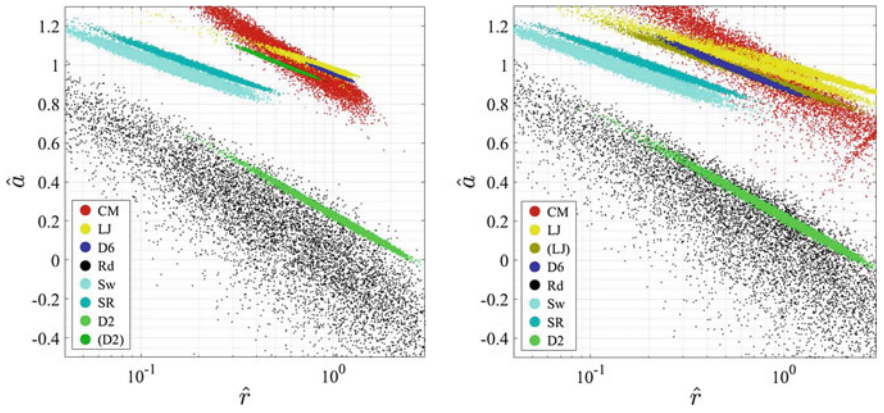

Fig. 13.2 Each dot in the figure corresponds to an individual epidemic that has been fit to Eq. (13.17). For each graph 10^4 epidemics were simulated. The overall "cloud" of point estimates of the parameters characterizes each graph in terms of what type of epidemics it produces. For the left figure maximum bound $c = 3$ was used and for the right figure Bernoulli thinning was used. Graph names in parenthesis indicate that the configuration model was used. Note the logarithmic scale on the horizontal axis

$\frac{J\tilde{\mu}}{n\mu}$, i.e. the proportion of already infected stubs divided by the total number of stubs. For the datasets we analyze this probability grows fastest for the LiveJournal dataset. This is because of the high quotient of $\frac{\tilde{\mu}}{\mu} \approx 5$. If we, arbitrarily, allow this probability to be at most 5%, thus reducing the instantaneous reproduction number by approximately the same amount, we cannot allow $\frac{J}{n}$ to be more than approximately 1% for the LiveJournal dataset and slightly higher for the other datasets. Setting this limit too low gives too few generations for the statistical analysis, thus increasing the confidence intervals for the parameters, and setting it too high means that the branching process approximation is no longer good and we should expect biased parameter estimates (generally too low values of a), indicating that the growth is not exponential), even when working with configuration model graphs. In this paper we set the limit to 1% for all datasets. We have tried (but not shown in this report) limits that are both lower and higher, and the chosen limit appeared to result in an acceptable compromise between imprecision and bias for the parameter estimates.

Finally, we make a couple of notes regarding the chosen model. First, we note that it is *not* a predictive model, but rather a way to characterize the initial phase of simulated epidemics on graphs. If we wanted to make predictions forward in time, then it is not certain that this model is the best method. We should then also validate the predictive properties of the model. Secondly, we are aware that data points are correlated, but we chose not to take this into account when fitting the data. We justify this by also including reference graphs in the same type of analysis that is used for the empirical graphs.

13.3 Results

In this section we present results of the statistical analysis for epidemics on the empirical graphs and compare the result with epidemics on some reference networks.

We have used 10^4 epidemics (with $I_0 = 1$) from each of the graphs in Table 13.1 and performed a least square fit to Eq. (13.17). Data used are from $i = 1$ until approximately a total of 1% of all vertices in the graph have been infected (see Sect. 13.2.5). For the maximum bound restriction of epidemics, we set $c = 3$ in the simulations. This is because some of the graphs do not allow for large epidemic outbreaks with $c = 2$, while $c = 3$ results in large epidemic on all graphs. For the corresponding simulation using Bernoulli thinning to restrict the epidemic, p was selected to give a *similar* growth rate early in the epidemic.

The results are summarized in Fig. 13.2 where we have plotted the estimated values of parameters a versus r. We remind the reader that a corresponds to the shape of growth where values close to 1 indicate exponential growth and values close to 0 indicate linear growth, while r is a measure of the growth rate, corresponding to the *Malthusian parameter* when we have exponential growth. In this paper the estimates of a are of most interest. For reference, we have included some configuration model and square lattice graphs together with the empirical graphs.

We note that most of the graphs produce epidemics with estimated parameter values in the vicinity of $a = 1$, while the road network and the square lattice data are spread out around $a = 0$. This indicates that most of the graphs produce epidemics that grow exponentially early on, while the road network and the square lattice show an essentially linear growth. Note the similarity between the configuration model simulation of D2 (the square lattice) and some of the empirical graphs. Somewhat surprising to the authors is that D6 (the six dimensional lattice) seems to produce restricted epidemics that grow exponentially, while we would have expected polynomial growth for these (in this case with $a = 4/5$, while the median estimated a-value is approximately 0.95). The explanation is that because of the relatively high dimension of the graph and the strong restriction on the spread of the epidemic, in the initial phase of the epidemic vertices still have many available neighbours that are not yet infected and the epidemic can be approximated by a branching process. While this would eventually change to polynomial growth if allowed to continue long enough, there is no space for this in a finite graph. Note that low r may also be a sign of non-exponential growth, since exponential growth with base close to 1 is hardly distinguishable from polynomial growth.

To better be able to observe differences in the estimated values of a, box plots of the a-estimates are shown in Fig. 13.3. In the plot outliers have been ignored to make the central part of the data more visible.

In order to see the effect of the restriction we place on the epidemic we show D2 and D6 using the maximum bound restriction, with different values of c, in Fig. 13.4. We note that lower values of c (more restricted epidemic) move the parameter estimates towards higher a-estimate for both lattices. For the D2 graph event when we use the highest possible restriction $c = 3$ the growth is still clearly polynomial, but for

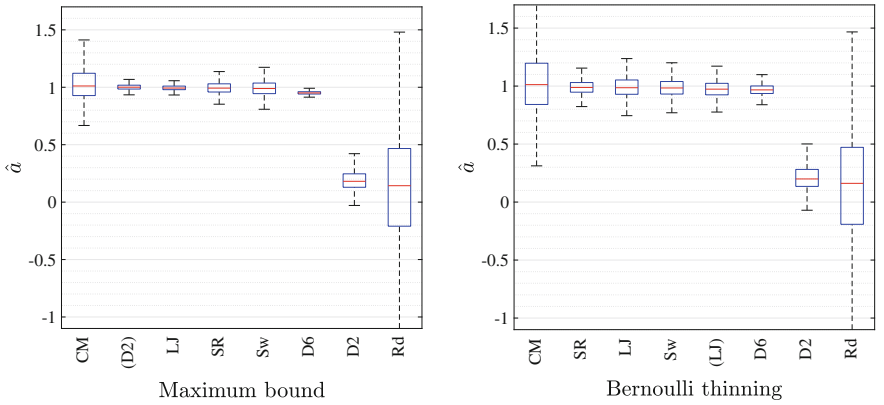

Fig. 13.3 The figure shows a box plot of the estimates of a for each graph from Fig. 13.2. Graph names in parenthesis indicate that the configuration model was used

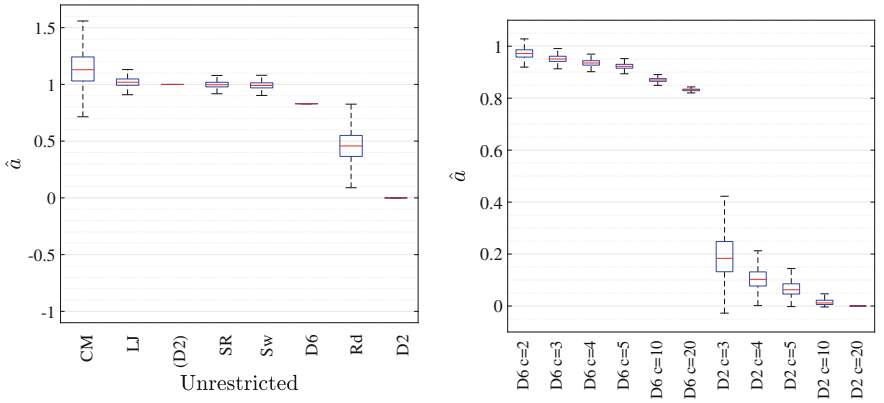

Fig. 13.4 The figure to the left shows a box plot for unrestricted epidemics. Graph names in parenthesis indicate that the configuration model was used. The figure to the right shows epidemics restricted with different values of c

the D6 graph we can shift the a-estimates so close to 1 that epidemics on the graph appear exponential. We conclude that if the epidemic spreads through only some smaller fraction of the available edges we can see exponential growth early on in the epidemic. One underlying assumption for this conclusion is that large epidemics are possible in the first place, i.e. that the graph is sufficiently well connected.

From the plots we also see that epidemics on the road network show much more variation than on the square lattice. The road network seems to be a mixture of strongly connected portions and long stretches of vertices in long lines connected only by single edges along the way. This is what we may expect from a road network for a large geographical area consisting both of densely connected cities and loosely connected countryside.

13.4 Discussion

The main purpose of this paper is to find a method to distinguish between empirical graphs which allow for initial exponential growth of an SIR epidemic and graphs which do not. If we know that exponential growth or close to exponential growth is possible, we can use statistical machinery already created for analyzing the growth potential of epidemics. To make this distinction we use the generalized growth model of [24] as presented in Sect. 13.2.5 above. This model has three parameters, but only a (which describes the shape—polynomial or exponential—of the initial growth) and r (which is a measure of the rate of growth) are relevant in this paper. We are mainly interested in a, but we cannot ignore r because the estimates of the two parameters are strongly dependent. Indeed, we see in Fig. 13.2 that although different epidemic simulations can produce very different parameter estimates, in the (r, a) plot estimates of epidemics on different underlying networks can still be distinguished.

Ideally, when a is close to 1 (how close depends on the application) we may conclude that the graph allows for epidemics that exhibit exponential growth. In (Fig. 13.3) we visualize the distribution of the estimates for parameter a for the individual graphs. This graph gives an indication of how close the growth of the epidemic is to exponential growth, but the figure must be interpreted with care. If the growth of the epidemic exactly follows the model with parameter $a = 1$ and r close to 0, then the growth is indeed exponential, but still very slow and it is very hard to distinguish this exponential growth from polynomial growth, with a larger r. Because the empirical networks are finite and we only observe a limited number of generations, we often do not have enough data to reliably distinguish between exponential growth with a small growth rate and polynomial growth. This observation is articulated in Fig. 13.2, where we see that some simulated epidemics on the road network (which is clearly two dimensional) produce estimates of a that are close to 1. However, for those simulations also the obtained estimates of r are low (typically 0.1 or lower).

For the Swedish population dataset comparing the original dataset with a randomized version indicates that there are some effects that may be attributed to spatial constraints, but the difference is mainly seen on the rate of growth through the parameter r and no so much on the parameter a. A possible conclusion is that the spatial constraints slow down the epidemic, but that the growth is still close to exponential (see e.g. [22] for a purely spatial model which allows for exponential growth of the epidemic).

The analysis of epidemics on the six dimensional lattice indicate that when the epidemic is restricted as in this report (Sect. 13.2.4) the resulting initial epidemic growth is essentially exponential. This can be explained as follows. Because vertices infect only a few of its neighbours, most neighbours of infected vertices will still be susceptible, so the local depletion of susceptibles is only felt after several generations, when probably already a considerable fraction of all the vertices are no longer susceptible. In addition, on an infinite six dimensional lattice I_n will grow as

a five dimensional polynomial, which corresponds with an a-value of $4/5$ in (13.16), which is relatively close to 1.

In the present work we only considered point estimates for the (r, a) parameter pair, in future work it is worth studying confidence regions for those parameters, based on one single observed epidemic on a network. In addition to summarizing data by fitting it to a model, the strength of models is to be able to make predictions. There are two classes of predictions we might desire. We may want to predict the continued development of a single epidemic in the future based on how it developed up until some point in time. We may also want to predict the development of future (new) epidemics on the same graph based on knowledge of a (limited) number of previous epidemics. For these predictions it is essential that we know whether we may expect exponential growth or not. We have not attempted to investigate the possibility of making such predictions in this paper, but it is certainly worth studying in future work.

Acknowledgements P.T. was supported by Vetenskapsrådet (Swedish Research Council), project 201604566. The authors would like to thank Tom Britton for valuable discussions.

References

1. Ball, F., Donnelly, P.: Strong approximations for epidemic models. Stoch. Process. Appl. **55**(1), 1–21 (1995). https://doi.org/10.1016/0304-4149(94)00034-Q
2. Ball, F., Sirl, D., Trapman, P.: Threshold behaviour and final outcome of an epidemic on a random network with household structure. Adv. Appl. Probab. **41**(3), 765–796 (2009)
3. Britton, T., Deijfen, M., Martin-Löf, A.: Generating simple random graphs with prescribed degree distribution. J. Stat. Phys. **124**(6), 1377–1397 (2006)
4. Diekmann, O., Heesterbeek, H., Britton, T.: Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press, Princeton (2012)
5. Durrett, R.: Random Graph Dynamics. Cambridge University Press, Cambridge (2006)
6. Erdős, P., Rényi, A.: On random graphs, i. Publ. Math. (Debrecen) **6**, 290–297 (1959)
7. Grassberger, P.: Two-dimensional SIR epidemics with long range infection. J. Stat. Phys. **153**(2), 289–311 (2013)
8. Grimmett, G.: Percolation, vol. 321, 2nd edn. Springer, Berlin (1999)
9. Guttorp, P.: Statistical Inference for Branching Processes, vol. 122. Wiley-Interscience, New York (1991)
10. van der Hofstad, R.: Random Graphs and Complex Networks, vol. 1. Cambridge University Press, Cambridge (2016)
11. Holm, E.: The SVERIGE spatial microsimulation model: content, validation, and example applications. Department of Social and Economic Geography, Umeå University (2002)
12. Jagers, P.: Branching Processes with Biological Applications. Wiley, New York (1975)
13. Leskovec, J., Krevl, A.: Snap datasets: Stanford large network dataset collection (2014). http://snap.stanford.edu/data
14. Malmros, J., Liljeros, F., Britton, T.: Respondent-driven sampling and an unusual epidemic. J. Appl. Probab. **53**(2), 518–530 (2016)
15. Mollison, D.: Spatial contact models for ecological and epidemic spread. J. R. Stat. Soc. Ser. B Stat. Methodol. **39**(3), 283–326 (1977)
16. Molloy, M., Reed, B.: A critical point for random graphs with a given degree sequence. Random Struct. Algorithms **6**(2–3), 161–180 (1995)

17. Newman, M.E.J.: Spread of epidemic disease on networks. Phys. Rev. E **66**(1), 016,128, 11 (2002). https://doi.org/10.1103/PhysRevE.66.016128
18. Reppell, M., Boehnke, M., Zöllner, S.: The impact of accelerating faster than exponential population growth on genetic variation. Genetics **196**(3), 819–828 (2014)
19. Tamhane, A.C., Dunlop, D.D.: Statistics and Data Analysis. Prentice Hall, Upper Saddle River (2000)
20. The MathWorks, Inc., Natick, Massachusetts, United States: MATLAB and curve fitting toolbox release (2017)
21. Tolle, J.: Can growth be faster than exponential, and just how slow is the logarithm?. Math. Gaz. **87**(510), 522–525 (2003). https://doi.org/10.1017/S0025557200173802
22. Trapman, P.: The growth of the infinite long-range percolation cluster. Ann. Probab. **38**(4), 1583–1608 (2010)
23. Trapman, P., Ball, F., Dhersin, J.S., Tran, V.C., Wallinga, J., Britton, T.: Inferring R_0 in emerging epidemics - the effect of common population structure is small. J. R. Soc. Interface **13**(121), 20160288, 9 pp (2016)
24. Viboud, C., Simonsen, L., Chowell, G.: A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. Epidemics **15**, 27–37 (2016)
25. Wallace, R.: Traveling waves of HIV infection on a low dimensional 'socio-geographic' network. Soc. Sci. Med. **32**(7), 847–852 (1991)
26. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440 (1998)

Chapter 14
Replication of Wiener-Transformable Stochastic Processes with Application to Financial Markets with Memory

Elena Boguslavskaya, Yuliya Mishura and Georgiy Shevchenko

Abstract We investigate Wiener-transformable markets, where the driving process is given by an adapted transformation of a Wiener process. This includes processes with long memory, like fractional Brownian motion and related processes, and, in general, Gaussian processes satisfying certain regularity conditions on their covariance functions. Our choice of markets is motivated by the well-known phenomena of the so-called "constant" and "variable depth" memory observed in real world price processes, for which fractional and multifractional models are the most adequate descriptions. Motivated by integral representation results in general Gaussian setting, we study the conditions under which random variables can be represented as pathwise integrals with respect to the driving process. From financial point of view, it means that we give the conditions of replication of contingent claims on such markets. As an application of our results, we consider the utility maximization problem in our specific setting. Note that the markets under consideration can be both arbitrage and arbitrage-free, and moreover, we give the representation results in terms of bounded strategies.

Keywords Wiener-transformable process · Fractional Brownian motion · Long memory · Pathwise integral · Martingale representation · Utility maximization

E. Boguslavskaya
Department of Mathematics, Brunel University London, Uxbridge UB8 3PH, UK
e-mail: elena@boguslavsky.net

Y. Mishura (✉) · G. Shevchenko
Department of Probability Theory, Statistics and Actuarial Mathematics, Taras Shevchenko National University of Kyiv, 64, Volodymyrs'ka St., Kyiv 01601, Ukraine
e-mail: myus@univ.kiev.ua

G. Shevchenko
e-mail: zhora@univ.kiev.ua

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_14

14.1 Introduction

Consider a general continuous time market model with one risky asset. For simplicity, we will work with discounted values. Let the stochastic process $\{X_t, t \in [0, T]\}$ model the discounted price of risky asset. Then the discounted final value of a self-financing portfolio is given by a stochastic integral

$$V^{\psi}(T) = V^{\psi}(0) + \int_0^T \psi(t) dX(t), \tag{14.1}$$

where an adapted process ψ is the quantity of risky asset in the portfolio. Loosely speaking, the self-financing assumption means that no capital is withdrawn or added to the portfolio; for precise definition and general overview of financial market models with continuous time we refer a reader to [4, 12].

Formula (14.1) raises several important questions of financial modeling, we will focus here on the following two.

- *Replication*:
  identifying random variables (i.e. discounted contingent claims), which can be represented as final capitals of some self-financing portfolios. In other words, one looks at integral representations

$$\xi = \int_0^T \psi(t) dX(t) \tag{14.2}$$

  with adapted integrand ψ; the initial value may be subtracted from ξ, so we can assume that it is zero.
- *Utility maximization*:
  maximizing the expected utility of final capital over some set of admissible self-financing portfolios.

An important issue is the meaning of stochastic integral in (14.1) or (14.2). When the process X is a semimartingale, it can be understood as Itô integral. In this case (14.1) is a kind of Itô representation, see e.g. [11] for an extensive coverage of this topic. When the Itô integral is understood in some extended sense, then the integral representation may exist under very mild assumptions and may be non-unique. For example, if $X = W$, a Wiener process, and ψ satisfies $\int_0^T \psi_s^2 ds < \infty$ a.s., then, as it was shown by [7], any random variable can be represented as a final value of some self-financing portfolio for any value of initial capital.

However, empirical studies suggest that financial markets often exhibit long-range dependence (in contrast to stochastic volatility that can be both smooth and rough, i.e., can demonstrate both long-and short-range dependence). The standard model for the phenomenon of long-range dependence is the fractional Brownian motion with Hurst index $H > 1/2$. It is not a semimartingale, so the usual Itô integration theory is not available. The standard approach now is to define the stochastic integral

in such models as a pathwise integral, namely, one usually considers the fractional integral, see [2, 23].

The models based on the fractional Brownian motion usually admit arbitrage possibilities, i.e. there are self-financing portfolios ψ such that $V_\psi(0) \le 0$, $V_\psi(T) \ge 0$ almost surely, and $V_\psi(T) > 0$ with positive probability. In the fractional Black–Scholes model, where $X_t = X_0 \exp\{at + bB_t^H\}$, and B^H is a fractional Brownian motion with $H > 1/2$, the existence of arbitrage was shown in [19]. Specifically, the strategy constructed there was of a "doubling" type, blowing the portfolio in the case of negative values; thus the potential intermediate losses could be arbitrarily large. It is worth to mention that such arbitrage exists even in the classical Black–Scholes model: the aforementioned result by Dudley allows gaining any positive final value of capital from initial zero by using a similar "doubling" strategy. For this reason, one usually restricts the class of admissible strategies by imposing a lower bound on the running value:

$$V^\psi(t) \ge -a, \quad t \in (0, T), \tag{14.3}$$

which in particular disallows the "doubling" strategies. However, in the fractional Black–Scholes model, the arbitrage exists even in the class of strategies satisfying (14.3), as was shown in [6].

There are several ways to exclude arbitrage in the fractional Brownian model. One possibility is to restrict the class of admissible strategies. For example, in [6] the absence of arbitrage is proved under further restriction that interval between subsequent trades is bounded from below (i.e. high frequency trading is prohibited). Another possibility is to add to the fractional Brownian motion an independent Wiener process, thus getting the so-called mixed fractional Brownian motion $M^H = B^H + W$. The absence in such mixed models was addressed in [1, 5]. In [1], it was shown that there is no arbitrage in the class of self-financing strategies $\gamma_t = f(t, M^H, t)$ of Markov type, depending only on the current value of the stock. In [5], it was shown that for $H \in (3/4, 1)$ the distribution of mixed fractional Brownian motion on a finite interval is equivalent to that of Wiener process. As a result, in such models there is no arbitrage strategies satisfying the non-doubling assumption (14.3). A more detailed exposition concerning arbitrage in models based on fractional Brownian motion is given in [3].

The replication question, i.e. the question when a random variable can be represented as a pathwise (fractional) integral in the models with long memory was studied in many articles, even in the case where arbitrage opportunities are present. The first results were established in [16], where it was shown that a random variable ξ has representation (14.2) with respect to fractional Brownian motion if it is a final value of some Hölder continuous adapted process. The assumption of Hölder continuity might seem too restrictive at the first glance. However, the article [16] gives numerous examples of random variables satisfying this assumption.

The results of [16] were extended in [22], where similar results were shown for a wide class of Gaussian integrators. The article [15] extended them even further and studied when a combination of Hölder continuity of integrator and small ball estimates lead to existence of representation (14.2).

For the mixed fractional Brownian motion, the question of replication was considered in [22]. The authors defined the integral with respect to fractional Brownian motion in pathwise sense and that with respect to Wiener process in the extended Itô sense and shown, similarly to the result of [7], that any random variable has representation (14.2).

It is worth to mention that the representations constructed in [15, 16, 22] involve integrands of "doubling" type, so in particular they do not satisfy the admissibility assumption (14.3).

Our starting point for this article was to see what contingent claims are representable as final values of some Hölder continuous adapted processes. It turned out that the situation is quite transparent whenever the Gaussian integrator generates the same flow of sigma-fields as the Wiener process. As a result, we came up with the concept of Wiener-transformable financial market, which turned out to be a fruitful idea, as a lot of models of financial markets are Wiener-transformable. We consider many examples of such models in our paper. Moreover, the novelty of the present results is that we prove representation theorems that, in financial interpretation, are equivalent to the possibility of hedging of contingent claims, in the class of *bounded* strategies. While even with such strategies the non-doubling assumption (14.3) may fail, the boundedness seems a feasible admissibility assumption.

More specifically, in the present paper we study a replication and the utility maximization problems for a broad class of asset prices processes, which are obtained by certain adapted transformation of a Wiener process; we call such processes *Wiener-transformable* and provide several examples. We concentrate mainly on non-semimartingale markets because the semimartingale markets have been studied thoroughly in the literature. Moreover, the novelty of the present results is that we prove representation theorems that, in financial interpretation, are equivalent to the possibility of hedging of contingent claims, in the class of bounded strategies. We would like to draw the attention of the reader once again to the fact that the possibility of representation means that we have arbitrage possibility in the considered class of strategies and they may be limited, although in a narrower and more familiar class of strategies the market can be arbitrage-free. Therefore, our results demonstrate rather subtle differences in the properties of markets in different classes of strategies.

The article is organized as follows. In Sect. 14.2, we recall basics of pathwise integrations in the fractional sense. In Sect. 14.3, we prove a new representation result, establishing an existence of integral representation with bounded integrand, which is of particular importance in financial applications. We also define the main object of study, Wiener-transformable markets, and provide several examples. Section 14.4 is devoted to application of representation results to the utility maximization problems.

14.2 Elements of Fractional Calculus

As announced in the introduction, the integral with respect to Wiener-transformable processes will be defined in pathwise sense, as fractional integral. Here we present the basic facts on fractional integration; for more details see [20, 23]. Consider functions

$f, g : [0, T] \to \mathbb{R}$, and let $[a, b] \subset [0, T]$. For $\alpha \in (0, 1)$ define Riemann-Liouville fractional derivatives

$$\big(\mathcal{D}^{\alpha}_{a+} f\big)(x) = \frac{1}{\Gamma(1-\alpha)}\left(\frac{f(x)}{(x-a)^{\alpha}} + \alpha \int_a^x \frac{f(x) - f(u)}{(x-u)^{1+\alpha}} du\right) 1_{(a,T)}(x),$$

$$\big(\mathcal{D}^{\alpha}_{b-} g\big)(x) = \frac{1}{\Gamma(1-\alpha)}\left(\frac{g(x)}{(b-x)^{\alpha}} + \alpha \int_x^b \frac{g(x) - g(u)}{(u-x)^{1+\alpha}} du\right) 1_{(0,b)}(x). \quad (14.4)$$

Assuming that $\mathcal{D}^{\alpha}_{a+} f \in L_1[a, b]$, $\mathcal{D}^{1-\alpha}_{b-} g_{b-} \in L_{\infty}[a, b]$, where $g_{b-}(x) = g(x) - g(b)$, the generalized Lebesgue–Stieltjes integral is defined as

$$\int_a^b f(x) dg(x) = \int_a^b \big(\mathcal{D}^{\alpha}_{a+} f\big)(x) \big(\mathcal{D}^{1-\alpha}_{b-} g_{b-}\big)(x) dx.$$

Let function g be θ-Hölder continuous, $g \in C^{\theta}[a, b]$ with $\theta \in (\frac{1}{2}, 1)$, i.e.

$$\sup_{t,s \in [0,T], t \neq s} \frac{|g(t) - g(s)|}{|t - s|^{\theta}} < \infty.$$

In order to integrate w.r.t. function g and to find an upper bound of the integral, fix some $\alpha \in (1 - \theta, 1/2)$ and introduce the following norm:

$$\|f\|_{\alpha,[a,b]} = \int_a^b \left(\frac{|f(s)|}{(s-a)^{\alpha}} + \int_a^s \frac{|f(s) - f(z)|}{(s-z)^{1+\alpha}} dz\right) ds.$$

For simplicity we abbreviate $\|\cdot\|_{\alpha,t} = \|\cdot\|_{\alpha,[0,t]}$. Denote

$$\Lambda_{\alpha}(g) := \sup_{0 \le s < t \le T} |\mathcal{D}^{1-\alpha}_{t-} g_{t-}(s)|.$$

In view of Hölder continuity, $\Lambda_{\alpha}(g) < \infty$.

Then for any $t \in (0, T]$ and for any f with $\|f\|_{\alpha,t} < \infty$, the integral $\int_0^t f(s) dg(s)$ is well defined as a generalized Lebesgue–Stieltjes integral, and the following bound is evident:

$$\left|\int_0^t f(s) dg(s)\right| \le \Lambda_{\alpha}(g) \|f\|_{\alpha,t}. \quad (14.5)$$

It is well known that in the case if f is β-Hölder continuous, $f \in C^{\beta}[a, b]$, with $\beta + \theta > 1$, the generalized Lebesgue–Stieltjes integral $\int_a^b f(x) dg(x)$ exists, equals to the limit of Riemann sums and admits bound (14.5) for any $\alpha \in (1 - \theta, \beta \wedge 1/2)$.

14.3 Representation Results for Gaussian and Wiener-Transformable Processes

Let throughout the paper $(\Omega, \mathcal{F}, \mathbf{P})$ be a complete probability space supporting all stochastic processes mentioned below. Let also $\mathbb{F} = \{\mathcal{F}_t, t \in [0, T]\}$ be a filtration satisfying standard assumptions. In what follows, the adaptedness of a process $X = \{X(t), t \in [0, T]\}$ will be understood with respect to $\mathbb{F}$, i.e. X will be called adapted if for any $t \in [0, T]$, $X(t)$ is $\mathcal{F}_t$-measurable.

We start with representation results, which supplement those of [15].

Consider a continuous centered Gaussian process G with incremental variance of G satisfying the following two-sided power bounds for some $H \in (1/2, 1)$.

(A) There exist $C_1, C_2 > 0$ such that for any $s, t \in [0, T]$

$$C_1 |t - s|^{2H} \leq \mathbf{E} |G(t) - G(s)|^2 \leq C_2 |t - s|^{2H}. \tag{14.6}$$

Assume additionally that the increments of G are positively correlated. More exactly, let the following condition hold

(B) For any $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2 \leq T$

$$\mathbf{E} (G(t_1) - G(s_1)) (G(t_2) - G(s_2)) \geq 0.$$

A process satisfying (14.6) is often referred to as a *quasi-helix*.

Note that the right inequality in (14.6) implies that

$$\sup_{t,s \in [0,T]} \frac{|G(t) - G(s)|}{|t - s|^H |\log(t - s)|^{1/2}} < \infty \tag{14.7}$$

almost surely (see e.g. p. 220 in [14]).

We will need the following small deviation estimate for sum of squares of Gaussian random variables, see e.g. [13].

Lemma 14.1 *Let $\{\xi_i\}_{i=1,\dots,n}$ be jointly Gaussian centered random variables. For all x such that $0 < x < \sum_{i=1}^n \mathbf{E}\xi_i^2$, it holds*

$$\mathbf{P}\left(\sum_{i=1}^n \xi_i^2 \leq x\right) \leq \exp\left\{-\frac{\left(\sum_{i=1}^n \mathbf{E}\xi_i^2 - x\right)^2}{\sum_{i,j=1}^n (\mathbf{E}\xi_i \xi_j)^2}\right\}.$$

Theorem 14.1 *Let a centered Gaussian process G satisfy (A) and (B) and ξ be a random variable such that there exists an adapted r-Hölder continuous process Z with $Z(T) = \xi$. There exists a bounded adapted process ψ, such that $\|\psi\|_{\alpha,T} < \infty$ for some $\alpha \in (1 - H, 1)$ and ξ admits the representation*

$$\xi = \int_0^T \psi(s) dG(s), \tag{14.8}$$

almost surely.

Remark 14.1 A similar result was proved in [15], Theorem 4.1, which assumed (14.6) with different exponents in the right-hand side and in the left-hand side of the inequality. Having equal exponents allowed us to establish existence of a *bounded* integrand ψ, thus extending previous results.

Proof To construct an integrand, we modify ideas of [15] and [21]. Throughout the proof, C will denote a generic constant, while $C(\omega)$, a random constant; their values may change between lines.

Choose some $\alpha \in \left(1 - H, (r + 1 - H) \wedge \frac{1}{2}\right)$.

We start with the construction of ψ. First take some $\theta \in (0, 1)$, put $t_n = T - \theta^n$, $n \geq 1$, and let $\Delta_n = t_{n+1} - t_n$. It is easy to see that

$$T - t_n \leq C\Delta_n. \tag{14.9}$$

Denote for brevity $\xi_n = Z(t_n)$. Then by Assumption 1, $|\xi_n - \xi_{n+1}| \leq C(\omega)\theta^{rn}$. Therefore, there exists some $N_0 = N_0(\omega)$ such that

$$|\xi_n - \xi_{n+1}| \leq n\theta^{rn} \tag{14.10}$$

for all $n \geq N_0(\omega)$.

We construct the integrand ψ inductively between the points $\{t_n, n \geq 1\}$. First let $\psi(t) = 0$, $t \in [0, t_1]$. Assuming that we have already constructed $\psi(t)$ on $[0, t_n)$, define $V(t) = \int_0^t \psi(s) dG(s)$, $t \in [0, t_n]$.

Consider some cases.

Case I. $V(t_n) \neq \xi_{n-1}$. By Lemma 4.1 in [15], there exists an adapted process $\{\phi_n(t), t \in [t_n, t_{n+1}]\}$, bounded on $[t_n, t]$ for any $t \in (t_n, t_{n+1})$ and such that $\int_{t_n}^t \phi_n(s) dG(s) \to +\infty$ as $t \to t_{n+1}-$. Define a stopping time

$$\tau_n = \inf\left\{t \geq t_n : \int_{t_n}^t \phi_n(s) dG(s) \geq |\xi_n - V(t_n)|\right\},$$

and set

$$\psi(t) = \phi_n(t) \operatorname{sign}\left(\xi_n - V(t_n)\right) 1_{[t_n, \tau_n]}(t),\ t \in [t_n, t_{n+1}).$$

It is obvious that $\int_{t_n}^{t_{n+1}} \psi(s) dG(s) = \xi_n - V(t_n)$ and $V(t_{n+1}) = \xi_n$.

Case II. $V(t_n) = \xi_{n-1}$. We consider a uniform partition $s_{n,k} = t_n + k\delta_n$, $k = 1, \ldots, n$ of $[t_n, t_{n+1}]$ with a mesh $\delta_n = \Delta_n / n$ and an auxiliary function

$$\phi_n(t) = a_n \sum_{k=0}^{n-1} \big(G(t) - G(s_{n,k})\big) 1_{[s_{n,k}, s_{n,k+1})}(t),$$

where $a_n = n^{-2}\theta^{(\alpha-H-1)n}$. Since ϕ_n is piecewise Hölder continuous of order up to H, by the change of variables formula (Theorem 4.3.1 in [23])

$$\int_{t_n}^{t_{n+1}} \phi_n(t) dG(t) = a_n \sum_{k=0}^{n-1} \big(G(s_{n,k+1}) - G(s_{n,k})\big)^2.$$

Define a stopping time

$$\sigma_n = \inf\left\{t \geq t_n : \int_{t_n}^{t} \phi_n(s) dG(s) \geq |\xi_n - \xi_{n-1}|\right\} \wedge t_{n+1},$$

and set

$$\psi(t) = \operatorname{sign}(\xi_n - \xi_{n-1}) \phi_n(t) 1_{[t_n, \sigma_n]}(t),\ t \in [t_n, t_{n+1}).$$

Now we want to ensure that, almost surely, $V(t_n) = \xi_{n-1}$ for all n large enough. By construction, Case I is always succeeded by Case II. So we need to ensure that $\sigma_n < t_{n+1}$ for all n large enough, equivalently, that

$$a_n \sum_{k=0}^{n-1} \big(G(s_{n,k+1}) - G(s_{n,k})\big)^2 > |\xi_n - \xi_{n-1}|.$$

Thanks to (14.10), it is enough to ensure that

$$\sum_{k=0}^{n-1} \big(G(s_{n,k+1}) - G(s_{n,k})\big)^2 > a_n^{-1} n\theta^{rn} = n^2\theta^{(r+H+1-\alpha)n}$$

for all n large enough. Define $\xi_k = G(s_{n,k+1}) - G(s_{n,k})$, $k = 0, \ldots, n-1$. Thanks to our choice of α, $r + H + 1 - \alpha > 2H$, so $n^2\theta^{(r+H+1-\alpha)n} < C_1 n^{1-2H}\theta^{2Hn}$ for all n large enough. Therefore, in view of (14.6),

$$\sum_{k=0}^{n-1} \mathbf{E}\xi_k^2 \geq C_1 n \delta_n^{2H} = C_1 n^{1-2H}\theta^{2Hn} > n^2\theta^{(r+H+1-\alpha)n},$$

so we can use Lemma 14.1. Using (A) and (B), estimate

$$\sum_{i,j=0}^{n-1}\big(\mathbf{E}\xi_i\xi_j\big)^2 \le \max_{0\le i,j\le n-1}\mathbf{E}\xi_i\xi_j\sum_{i,j=0}^{n-1}\mathbf{E}\xi_i\xi_j$$

$$\le C_1\delta_n^{2H}\mathbf{E}\Big(\sum_{i=0}^{n-1}\xi_i\Big)^2 = C_1\delta_n^{2H}\mathbf{E}\big(G(t_{n+1})-G(t_n)\big)^2$$

$$\le C_1^2\delta_n^{2H}\Delta_n^{2H} \le C_1^2 n^{-2H}\Delta^{4H} = C_1^2 n^{-2H}\theta^{4Hn}.$$

Hence, by Lemma 14.1,

$$\mathbf{P}\left(\sum_{k=0}^{n-1}\big(G(s_{n,k+1})-G(s_{n,k})\big)^2 \le n^2\theta^{(r+H+1-\alpha)n}\right)$$

$$\le \exp\left\{-\frac{\big(C_1 n^{1-2H}\theta^{2Hn}-n^2\theta^{(r+H+1-\alpha)n}\big)^2}{C_1^2 n^{-2H}\theta^{4Hn}}\right\} \le \exp\left\{-Cn^{2-2H}\right\}.$$

Therefore, by the Borel–Cantelli lemma, almost surely there exists some $N_1(\omega)\ge N_0(\omega)$ such that for all $n\ge N_1(\omega)$

$$\sum_{k=0}^{n-1}\big(G(s_{n,k+1})-G(s_{n,k})\big)^2 > n^2\theta^{(r+H+1-\alpha)n},$$

so, as it was explained above, we have $V(t_n)=\xi_{n-1}$, $n\ge N_1(\omega)$.

Since all functions ϕ_n are bounded, we have that ψ is bounded on $[0,t_N]$ for any $N\ge 1$. Further, thanks to (14.7), for $t\in[t_n,t_{n+1}]$ with $n\ge N_1(\omega)$,

$$\begin{aligned}|\psi(s)| &\le C(\omega)a_n\delta_n^H\,|\log\delta_n|^{1/2} \le C(\omega)n^{-2}\theta^{(\alpha-H-1)n}n^{-H}\theta^{Hn}n^{1/2}\\ &= C(\omega)n^{\alpha-H-3/2}\theta^{(\alpha-1)n}.\end{aligned} \tag{14.11}$$

Therefore, ψ is bounded (moreover, $\psi(t)\to 0$, $t\to T-$).

Further, by construction, $\|\psi\|_{\alpha,t_N}<\infty$ for any $N\ge 1$. Moreover, $|V(t)-\xi_{N-1}|\le|\xi_N-\xi_{N-1}|$, $t\in[t_N,t_{N+1}]$. Thus, it remains to to verify that $\|\psi\|_{\alpha,[t_N,1]}<\infty$ and $\int_{t_N}^1\psi(s)dG(s)\to 0$, $N\to\infty$, which would follow from $\|\psi\|_{\alpha,[t_N,1]}\to 0$, $N\to\infty$.

Let $N\ge N_1(\omega)$. Write

$$\|\psi\|_{\alpha,[t_N,T]} = \sum_{n=N}^{\infty}\int_{t_n}^{t_{n+1}}\left(\frac{|\psi(s)|}{(s-t_N)^\alpha}+\int_{t_N}^{s}\frac{|\psi(s)-\psi(u)|}{|s-u|^{1+\alpha}}du\right)ds.$$

Thanks to (14.11),

$$\int_{t_n}^{t_{n+1}}\frac{|\psi(s)|}{(s-t_N)^\alpha}ds \le C(\omega)\Delta_n^{1-\alpha}n^{\alpha-H-3/2}\theta^{(\alpha-1)n} = C(\omega)n^{\alpha-H-3/2}.$$

Further,

$$\int_{t_N}^{t_{n+1}} \int_{t_n}^{s} \frac{|\psi(s)-\psi(u)|}{|s-u|^{1+\alpha}} du\, ds$$
$$= \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} \left(\int_{t_N}^{t_n} + \int_{t_n}^{s_{n,k-1}} + \int_{s_{n,k-1}}^{s} \right) \frac{|\psi(s)-\psi(u)|}{|s-u|^{1+\alpha}} du\, ds =: I_1 + I_2 + I_3.$$

Start with I_1, observing that ψ vanishes on $(\sigma_n, t_{n+1}]$:

$$I_1 \le \int_{t_n}^{t_{n+1}} \sum_{j=N}^{n} \int_{t_{j-1}}^{t_j} \frac{|\psi(s)|+|\psi(u)|}{|s-u|^{1+\alpha}} du\, ds$$
$$\le C(\omega) n^{\alpha-H-3/2} \theta^{(\alpha-1)n} \int_{t_n}^{t_{n+1}} (s-t_n)^{-\alpha} ds$$
$$+ C(\omega) \sum_{j=N}^{n-1} j^{\alpha-H-3/2} \theta^{(\alpha-1)j} \int_{t_n}^{t_{n+1}} (s-t_{j+1})^{-\alpha} ds$$
$$\le C(\omega) n^{\alpha-H-3/2} \theta^{(\alpha-1)n} \Delta_n^{1-\alpha} + C(\omega) \sum_{j=N}^{n-1} j^{\alpha-H-3/2} \theta^{(\alpha-1)j} \Delta_n^{1-\alpha}$$
$$= C(\omega) n^{\alpha-H-3/2} + C(\omega) \sum_{j=N}^{n-1} j^{\alpha-H-3/2} \theta^{(\alpha-1)(j-n)}.$$

Similarly,

$$I_2 \le C(\omega) n^{\alpha-H-3/2} \theta^{(\alpha-1)n} \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} \int_{t_n}^{s_{n,k-1}} |s-u|^{-1-\alpha} du\, ds$$
$$\le C(\omega) n^{\alpha-H-3/2} \theta^{(\alpha-1)n} \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} (s-s_{n,k-1})^{-\alpha} ds$$
$$\le C(\omega) n^{\alpha-H-3/2} \theta^{(\alpha-1)n} n \delta_n^{1-\alpha} = C(\omega) n^{2\alpha-H-3/2}.$$

Finally, assuming that $\sigma_n \in [s_{n,l-1}, s_{n,l})$,

$$I_3 \le C(\omega)\sum_{k=1}^{l-1}\int_{s_{n,k-1}}^{s_{n,k}}\int_{s_{n,k-1}}^{s} a_n\frac{(s-u)^H|\log(s-u)|^{1/2}}{(s-u)^{1+\alpha}}du\,ds$$
$$+\int_{s_{n,l-1}}^{\sigma_n}\int_{s_{n,l-1}}^{s}\frac{|\psi(s)-\psi(u)|}{|s-u|^{1+\alpha}}du\,ds+\int_{\sigma_n}^{s_{n,l}}\int_{s_{n,l-1}}^{\sigma_n}\frac{|\psi(s)-\psi(u)|}{|s-u|^{1+\alpha}}du\,ds$$
$$\le C(\omega)a_n\sum_{k=1}^{n}\int_{s_{n,k-1}}^{s_{n,k}}(s-s_{n,k-1})^{H-\alpha}|\log(s-s_{n,k-1})|^{1/2}ds$$
$$+C(\omega)n^{\alpha-H-3/2}\theta^{(\alpha-1)n}\int_{\sigma_n}^{s_{n,l}}\int_{s_{n,l-1}}^{\sigma_n}\frac{1}{|s-u|^{1+\alpha}}du\,ds$$
$$\le C(\omega)a_n n\delta_n^{H+1-\alpha}|\log\delta_n|^{1/2}+C(\omega)n^{\alpha-H-3/2}\theta^{(\alpha-1)n}\delta_n^{-\alpha}$$
$$=C(\omega)n^{\alpha-H-3/2}+C(\omega)n^{2\alpha-H-3/2}\le C(\omega)n^{2\alpha-H-3/2}.$$

Gathering all estimates we get

$$\int_{t_N}^{1}|D^{\alpha}_{t_N+}(\psi)(s)|ds\le C(\omega)\sum_{n=N}^{\infty}\left(n^{2\alpha-H-3/2}+\sum_{j=N}^{n-1}j^{\alpha-H-3/2}\theta^{(\alpha-1)(j-n)}\right)$$
$$\le C(\omega)\left(N^{2\alpha-H-1/2}+\sum_{j=N}^{\infty}j^{\alpha-H-3/2}\sum_{n=j+1}^{\infty}\theta^{(1-\alpha)(n-j)}\right)$$
$$\le C(\omega)N^{2\alpha-H-1/2},$$

which implies that $\|\psi\|_{\alpha,[t_N,T]}\to 0$, $N\to\infty$, finishing the proof. □

Now we turn to the main object of this article.

Definition 14.1 A Gaussian process $G=\{G(t),t\in\mathbb{R}^+\}$ is called m-Wiener-transformable if there exists m-dimensional Wiener process $W=\{W(t),t\in\mathbb{R}^+\}$ such that G and W generate the same filtration, i.e. for any $t\in\mathbb{R}^+$

$$\mathcal{F}_t^G=\mathcal{F}_t^W.$$

We say that G is m-Wiener-transformable to W (evidently, process W can be non-unique.)

Remark 14.2 (i) In the case when $m=1$ we say that the process G is Wiener-transformable.

(ii) Being Gaussian so having moments of any order, m-Wiener-transformable process admits at each time $t\in\mathbb{R}^+$ the martingale representation $G(t)=\mathbf{E}(G(0))+\sum_{i=1}^m\int_0^t K_i(t,s)dW_i(s)$, where $K_i(t,s)$ is $\mathcal{F}_s^W$-measurable for any $0\le s\le t$ and $\int_0^t\mathbf{E}(K_i(t,s))^2ds<\infty$ for any $t\in\mathbb{R}^+$.

Now let the random variable ξ be $\mathcal{F}_T^W$-measurable, $\mathbf{E}\xi^2<\infty$. Then in view of martingale representation theorem, ξ can be represented as

$$\xi = \mathbf{E}\xi + \int_0^T \vartheta(t) dW(t), \tag{14.12}$$

where ϑ is an adapted process with $\int_0^T \mathbf{E}\vartheta(t)^2 dt < \infty$.

As it was explained in introduction, we are interested when ξ can be represented in the form

$$\xi = \int_0^T \psi(s) dG(s),$$

where the integrand is adapted, and the integral is understood in the pathwise sense.

Theorem 14.2 *Let the following conditions hold.*

(i) Gaussian process G satisfies condition (A) and (B).
(ii) Stochastic process ϑ in representation (14.12) satisfies

$$\int_0^T |\vartheta(s)|^{2p} ds < \infty \tag{14.13}$$

a.s. with some $p > 1$.

Then there exists a bounded adapted process ψ such that $\|\psi\|_{\alpha,T} < \infty$ for some $\alpha \in \left(1 - H, \frac{1}{2}\right)$ and ξ admits the representation

$$\xi = \int_0^T \psi(s) dG(s),$$

almost surely.

Remark 14.3 As it was mentioned in [15], it is sufficient to require the properties (A) and (B) to hold on some subinterval $[T - \delta, T]$. Similarly, it is enough to require in (ii) that $\int_{T-\delta}^T |\vartheta(t)|^{2p} dt < \infty$ almost surely.

First we prove a simple result establishing Hölder continuity of Itô integral.

Lemma 14.2 *Let $\vartheta = \{\vartheta(t), t \in [0, T]\}$ be a real-valued progressively measurable process such that for some $p \in (1, +\infty]$*

$$\int_0^T |\vartheta(s)|^{2p} ds < \infty$$

a.s. Then the stochastic integral $\int_0^t \vartheta(s) dW(s)$ is Hölder continuous of any order up to $\frac{1}{2} - \frac{1}{2p}$.

Proof First note that if there exist non-random positive constants a, C such that for any $s, t \in [0, T]$ with $s < t$

$$\int_s^t \vartheta^2(u)du \le C(t-s)^a,$$

then $\int_0^t \vartheta(s)dW(s)$ is Hölder continuous of any order up to $a/2$. Indeed, in this case by the Burkholder inequality, for any $r > 1$ and $s, t \in [0, T]$ with $s < t$

$$\mathbf{E}\left|\int_s^t \vartheta(u)dW(u)\right|^r \le C_r\mathbf{E}\left(\int_s^t \vartheta^2(u)du\right)^{r/2} \le C(t-s)^{ar/2},$$

so by the Kolmogorov–Chentsov theorem, $\int_0^t \vartheta(s)dW(s)$ is Hölder continuous of order $\frac{1}{r}(\frac{ar}{2}-1) = \frac{a}{2} - \frac{1}{2r}$. Since r can be arbitrarily large, we deduce the claim.

Now let for $n \ge 1$, $\vartheta_n(t) = \vartheta(t)\mathbf{1}_{\int_0^t |\vartheta(s)|^{2p}ds \le n}$, $t \in [0, T]$. By the Hölder inequality, for any $s, t \in [0, T]$ with $s < t$

$$\int_s^t \vartheta_n^2(u)du \le (t-s)^{1-1/p}\left(\int_s^t |\vartheta(u)|^{2p}du\right)^{1/p} \le n^{1/p}(t-s)^{1-1/p}.$$

Therefore, by the above claim, $\int_0^t \vartheta_n(s)dW(s)$ is a.s. Hölder continuous of any order up to $\frac{1}{2} - \frac{1}{2p}$. However, ϑ_n coincides with ϑ on $\Omega_n = \{\int_0^T |\vartheta(t)|^{2p}dt \le n\}$. Consequently, $\int_0^t \vartheta_n(s)dW(s)$ is a.s. Hölder continuous of any order up to $\frac{1}{2} - \frac{1}{2p}$ on Ω_n. Since $\mathbf{P}(\bigcup_{n\ge1}\Omega_n) = 1$, we arrive at the statement of the lemma. □

Proof of Theorem 14.2. Define

$$Z(t) = \mathbf{E}\xi + \int_0^t \vartheta(s)dW(s).$$

This is an adapted process with $Z(T) = \xi$, moreover, it follows from Lemma 14.2 that Z is Hölder continuous of any order up to $\frac{1}{2} - \frac{1}{2p}$. Thus, the statement follows from Theorem 14.1.

In the case where one looks at improper representation, no assumptions on ξ are needed.

Theorem 14.3 (Improper representation theorem) *Assume that an adapted Gaussian process $G = \{G(t),\ t \in [0, T]\}$ satisfies conditions (A), (B). Then for any random variable ξ there exists an adapted process ψ that $\|\psi\|_{\alpha,t} < \infty$ for some $\alpha \in \left(1 - H, \frac{1}{2}\right)$ and any $t \in [0, T)$ and ξ admits the representation*

$$\xi = \lim_{t\to T-}\int_0^t \psi(s)dG(s),$$

almost surely.

Proof The proof is exactly the same as for Theorem 4.2 in [22], so we just sketch the main idea.

Consider an increasing sequence of points $\{t_n, n \geq 1\}$ in $[0, T)$ such that $t_n \to T$, $n \to \infty$, and let $\{\xi_n, n \geq 1\}$ be a sequence of random variables such that ξ_n is $\mathcal{F}_{t_n}$-measurable for each $n \geq 1$, and $\xi_n \to \xi$, $n \to \infty$, a.s. Set for convenience $\xi_0 = 0$. Similarly to Case I in Theorem 14.1, for each $n \geq 1$, there exists an adapted process $\{\phi_n(t), t \in [t_n, t_{n+1}]\}$, such that $\int_{t_n}^{t} \phi_n(s) dG(s) \to +\infty$ as $t \to t_{n+1}-$. For $n \geq 1$, define a stopping time

$$\tau_n = \inf \left\{ t \geq t_n : \int_{t_n}^{t} \phi_n(s) dG(s) \geq |\xi_n - \xi_{n-1}| \right\}$$

and set

$$\psi(t) = \phi_n(t) \operatorname{sign} \left(\xi_n - \xi_{n-1}\right) 1_{[t_n, \tau_n]}(t), \ t \in [t_n, t_{n+1}).$$

Then for any $n \geq 1$, we have $\int_0^{t_{n+1}} \psi(s) dG(s) = \xi_n$ and $\int_0^t \psi(s) dG(s)$ lies between ξ_{n-1} and ξ_n for $t \in [t_{n-1}, t_n]$. Consequently, $\int_0^t \psi(s) dG(s) \to \xi$, $t \to T-$, a.s., as required. □

Further we give several examples of Wiener-transformable Gaussian processes satisfying conditions (A) and (B) (for more detail and proofs see, e.g. [15]) and formulate the corresponding representation results.

14.3.1 Fractional Brownian Motion

Fractional Brownian motion B^H with Hurst parameter $H \in (0, 1)$ is a centered Gaussian process with the covariance

$$\mathbf{E} B^H(t) B^H(s) = \frac{1}{2} \left(t^{2H} + s^{2H} - |t - s|^{2H} \right);$$

an extensive treatment of fractional Brownian motion is given in [17]. For $H = \frac{1}{2}$, fractional Brownian motion is a Wiener process; for $H \neq \frac{1}{2}$ it is Wiener-transformable to the Wiener process W via relations

$$B^H(t) = \int_0^t K^H(t, s) dW(s) \tag{14.14}$$

and

$$W(t) = \int_0^t k^H(t, s) dB^H(s), \tag{14.15}$$

see e.g. [18].

Fractional Brownian motion with index $H \in (0, 1)$ satisfies condition (A) and satisfies condition (B) if $H \in (\frac{1}{2}, 1)$.

Therefore, a random variable satisfying (14.13) with any $p > 1$ admits the representation (14.8).

14.3.2 Fractional Ornstein–Uhlenbeck Process

Let $H \in (\frac{1}{2}, 1)$. Then the fractional Ornstein–Uhlenbeck process $Y = \{Y(t), t \geq 0\}$, involving fractional Brownian component and satisfying the equation

$$Y(t) = Y_0 + \int_0^t (b - aY(s))ds + \sigma B^H(t),$$

where $a, b \in \mathbb{R}$ and $\sigma > 0$, is Wiener-transformable to the same Wiener process as the underlying fBm B^H.

Consider a fractional Ornstein–Uhlenbeck process of the simplified form

$$Y(t) = Y_0 + a \int_0^t Y(s)ds + B^H(t), \ t \geq 0.$$

It satisfies condition (A); if $a > 0$, it satisfies condition (B) as well.

As it was mentioned in [15], the representation theorem is valid for a fractional Ornstein-Uhlenbeck process with a negative drift coefficient too. Indeed, we can annihilate the drift of the fractional Ornstein-Uhlenbeck process with the help of Girsanov theorem, transforming a fractional Ornstein-Uhlenbeck process with negative drift to a fractional Brownian motion $\widetilde{B}^H$. Then, assuming (14.13), we represent the random variable ξ as $\xi = \int_0^T \psi(s)d\widetilde{B}^H(s)$ on the new probability space. Finally, we return to the original probability space. Due to the pathwise nature of integral, its value is not changed upon changes of measure.

14.3.3 Subfractional Brownian Motion

Subfractional Brownian motion with index H, that is a centered Gaussian process $G^H = \left\{G^H(t), t \geq 0\right\}$ with covariance function

$$\mathbf{E}G^H(t)G^H(s) = t^{2H} + s^{2H} - \frac{1}{2}\left(|t+s|^{2H} + |t-s|^{2H}\right),$$

satisfies condition (A) and condition (B) for $H \in (\frac{1}{2}, 1)$.

14.3.4 Bifractional Brownian Motion

Bifractional Brownian motion with indices $A \in (0, 1)$ and $K \in (0, 1)$, that is a centered Gaussian process with covariance function

$$\mathbf{E}G^{A,K}(t)G^{A,K}(s) = \frac{1}{2^K}\left(\left(t^{2A} + s^{2A}\right)^K - |t-s|^{2AK}\right),$$

satisfies condition (A) with $H = AK$ and satisfies condition (B) for $AK > \frac{1}{2}$.

14.3.5 Geometric Brownian Motion

Geometric Brownian motion involving the Wiener component and having the form

$$S = \{S(t) = S(0)\exp\{\mu t + \sigma W(t)\}, \quad t \geq 0\},$$

with $S(0) > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$, is Wiener-transformable to the underlying Wiener process W. However, it does not satisfy the assumptions of Theorem 14.2. One should appeal here to the standard semimartingale tools, like the martingale representation theorem.

14.3.6 Linear Combination of Fractional Brownian Motions

Consider a collection of Hurst indices $\frac{1}{2} \leq H_1 < H_2 < \ldots < H_m < 1$ and independent fractional Brownian motions with corresponding Hurst indices H_i, $1 \leq i \leq m$. Then the linear combination $\sum_{i=1}^m a_i B^{H_i}$ is m-Wiener-transformable to the Wiener process $W = (W_1, \ldots, W_m)$, where W_i is such Wiener process to which fractional Brownian motion B^{H_i} is Wiener-transformable. In particular, the mixed fractional Brownian motion $M^H = W + B^H$, introduced in [5], is 2-Wiener-transformable.

The linear combination $\sum_{i=1}^m a_i B^{H_i}$ satisfies condition (A) with $H = H_1$, and condition (B) whenever $H_1 > 1/2$.

We note that in the case of mixed fractional Brownian motion, the existence of representation (14.8) cannot be derived from Theorem 14.2, as we have $H = \frac{1}{2}$ in this case. By slightly different methods, it was established in [22] that arbitrary $\mathcal{F}_T$-measurable random variable ξ admits the representation

$$\xi = \int_0^T \psi(s)d\big(B^H(s) + W(s)\big),$$

where the integral with respect to B^H is understood, as here, in the pathwise sense, the integral with respect to W, in the extended Itô sense. In contrast to Theorem 14.1,

we can not for the moment establish this result for the bounded strategies. Therefore, it would be interesting to study which random variables have representations with bounded ψ in the mixed model.

14.3.7 Volterra Process

Consider Volterra integral transform of Wiener process, that is the process of the form $G(t) = \int_0^t K(t,s)dW(s)$ with non-random kernel $K(t,\cdot) \in L_2[0,t]$ for $t \in [0,T]$. Let the constant $r \in [0, 1/2)$ be fixed. Let the following conditions hold.

($B1$) The kernel K is non-negative on $[0,T]^2$ and for any $s \in [0,T]$ $K(\cdot,s)$ is non-decreasing in the first argument;

($B2$) There exist constants $D_i > 0$, $i = 2, 3$ and $H \in (1/2, 1)$ such that

$$|K(t_2,s) - K(t_1,s)| \le D_2|t_2 - t_1|^H s^{-r}, \quad s, t_1, t_2 \in [0,T]$$

and

$$K(t,s) \le D_3(t-s)^{H-1/2}s^{-r};$$

and at least one of the following conditions

($B3$,a) There exist constant $D_1 > 0$ such that

$$D_1|t_2 - t_1|^H s^{-r} \le |K(t_2,s) - K(t_1,s)|, \quad s, t_1, t_2 \in [0,T];$$

($B3$,b) There exist constant $D_1 > 0$ such that

$$K(t,s) \ge D_1(t-s)^{H-1/2}s^{-r}, \quad s, t \in [0,T].$$

Then the Gaussian process $G(t) = \int_0^t K(t,s)dW(s)$, satisfies condition (A), (B) on any subinterval $[T-\delta, T]$ with $\delta \in (0,1)$.

14.4 Expected Utility Maximization in Wiener-Transformable Markets

14.4.1 Expected Utility Maximization for Unrestricted Capital Profiles

Consider the problem of maximizing the expected utility. Our goal is to characterize the optimal asset profiles in the framework of the markets with risky assets involving Gaussian processes satisfying conditions of Theorem 14.2. We follow the general

approach described in [9, 12], but apply its interpretation from [10]. We fix $T > 0$ and from now on consider $\mathcal{F}_T^W$-measurable random variables. Let the utility function $u : \mathbb{R} \to \mathbb{R}$ be strictly increasing and strictly concave, $L^0(\Omega, \mathcal{F}_T^W, \mathbf{P})$ be the set of all $\mathcal{F}_T^W$-measurable random variables, and let the set of admissible capital profiles coincide with $L^0(\Omega, \mathcal{F}_T^W, \mathbf{P})$. Let $\mathbf{P}^*$ be a probability measure on $(\Omega, \mathcal{F}_T^W)$, which is equivalent to $\mathbf{P}$, and denote $\varphi(T) = \frac{d\mathbf{P}^*}{d\mathbf{P}}$. The budget constraint is given by $\mathbf{E}_{\mathbf{P}^*}(X) = w$, where $w > 0$ is some number that can be in some cases, but not obligatory, interpreted as the initial wealth. Thus the budget set is defined as

$$\mathcal{B} = \left\{ X \in L^0\left(\Omega, \mathcal{F}_T^W, \mathbf{P}\right) \cap L^1\left(\Omega, \mathcal{F}_T^W, \mathbf{P}^*\right) | \mathbf{E}_{\mathbf{P}^*}(X) = w \right\}.$$

The problem is to find such $X^* \in \mathcal{B}$, for which $\mathbf{E}(u(X^*)) = \max_{X \in \mathcal{B}} \mathbf{E}(u(X))$. Consider the inverse function $I(x) = (u'(x))^{-1}$.

Theorem 14.4 ([10], Theorem 3.34) *Let the following condition hold: Strictly increasing and strictly concave utility function $u : \mathbb{R} \to \mathbb{R}$ is continuously differentiable, bounded from above and*

$$\lim_{x \downarrow -\infty} u'(x) = +\infty.$$

Then the solution of this maximization problem has a form

$$X^* = I(c\varphi(T)),$$

under additional assumption that $\mathbf{E}_{\mathbf{P}^*}(X^*) = w$.

To connect the solution of maximization problem with specific W-transformable Gaussian process describing the price process, we consider the following items.

1. Consider random variable $\varphi(T)$, $\varphi(T) > 0$ a.s. and let $\mathbf{E}(\varphi(T)) = 1$. Being the terminal value of a positive martingale $\varphi = \{\varphi_t = \mathbf{E}(\varphi(T)|\mathcal{F}_t^W), t \in [0, T]\}$, $\varphi(T)$ admits the following representation

$$\varphi(T) = \exp\left\{ \int_0^T \vartheta(s) dW_s - \frac{1}{2} \int_0^T \vartheta^2(s) ds \right\}, \tag{14.16}$$

where ϑ is a real-valued progressively measurable process for which

$$\mathbf{P}\left\{ \int_0^T \vartheta^2(s) ds < \infty \right\} = 1.$$

Assume that ϑ satisfies (14.13). Then $\varphi(T)$ is a terminal value of a Hölder continuous process of order $\frac{1}{2} - \frac{1}{2p}$.

2. Consider W-transformable Gaussian process $G = \{G(t), t \in [0, T]\}$ satisfying conditions (A) and (B), and introduce the set

$$\mathcal{B}_w^G = \left\{ \psi : [0,T] \times \Omega \to \mathbb{R} \,\middle|\, \psi \text{ is bounded } \mathcal{F}_t^W - \text{adapted,} \right.$$

there exists a generalized Lebesgue- Stieltjes integral $\int_0^T \psi(s)dG(s)$ and

$$\left. \mathbf{E}\left(\varphi(T)\int_0^T \psi(s)dG(s)\right) = w \right\}.$$

Theorem 14.5 *Let the following conditions hold*

(i) Gaussian process G satisfies condition (A) and (B).
(ii) Function $I(x)$, $x \in \mathbb{R}$ is Hölder continuous.
(iii) Stochastic process ϑ in representation (14.16) satisfies (14.13) with some $p > 1$.
(iv) There exists $c \in \mathbb{R}$ such that $\mathbf{E}(\varphi(T)I(c\varphi(T))) = w$.

Then the random variable $X^* = I(c\varphi(T))$ admits the representation

$$X^* = \int_0^T \overline{\psi}(s)dG(s), \tag{14.17}$$

with some $\overline{\psi} \in \mathcal{B}_w^G$, and

$$\mathbf{E}(u(X^*)) = \max_{\psi \in \mathcal{B}_w^G} \mathbf{E}\left(u\left(\int_0^T \psi(s)dG(s)\right)\right). \tag{14.18}$$

Proof From Lemma 14.2 we have that for any $c \in \mathbb{R}$ the random variable $\xi = I(c\varphi(T))$ is the final value of a Hölder continuous process

$$U(t) = I(c\varphi(t)) = I\left(c\exp\left\{\int_0^t \vartheta(s)dW(s) - \frac{1}{2}\int_0^t \vartheta^2(s)ds\right\}\right),$$

and the Hölder exponent exceeds ρ. Together with (i)–(iii) this allows to apply Theorem 14.2 to obtain the existence of representation (14.17). Assume now that (14.18) is not valid, and there exists $\psi_0 \in \mathcal{B}_w^G$ such that $\mathbf{E}\left(\varphi(T)\int_0^T \psi_0(s)dG(s)\right) = w$, and $\mathbf{E}u\left(\int_0^T \psi_0(s)dG(s)\right) > \mathbf{E}u(X^*)$. But in this case $\int_0^T \psi_0(s)dG(s)$ belongs to $\mathcal{B}$, and we get a contradiction with Theorem 14.4. □

Remark 14.4 Assuming only (i) and (iv), one can show in a similar way, but using Theorem 14.3 instead of Theorem 14.2, that

$$\mathbf{E}(u(X^*)) = \sup_{\psi \in \mathcal{B}_w^G} \mathbf{E}\left(u\left(\int_0^T \psi(s)dG(s)\right)\right).$$

However, the existence of a maximizer is not guaranteed in this case.

Example 14.1 Let $u(x) = 1 - e^{-\beta x}$ be an exponential utility function with constant absolute risk aversion $\beta > 0$. In this case $I(x) = -\frac{1}{\beta}\log(\frac{x}{\beta})$. Assume that

$$\varphi(T) = \exp\left\{\int_0^T \vartheta(s)dW(s) - \frac{1}{2}\int_0^T \vartheta^2(s)ds\right\}$$

is chosen in such a way that

$$\begin{aligned}&\mathbf{E}\left(\varphi(T)|\log\varphi(T)|\right)\\ &= \mathbf{E}\Bigg(\exp\left\{\int_0^T \vartheta(s)dW(s) - \frac{1}{2}\int_0^T \vartheta^2(s)ds\right\}\\ &\times\left|\int_0^T \vartheta(s)dW(s) - \frac{1}{2}\int_0^T \vartheta^2(s)ds\right|\Bigg) < \infty.\end{aligned} \tag{14.19}$$

Then, according to Example 3.35 from [10], the optimal profile can be written as

$$X^* = -\frac{1}{\beta}\left(\int_0^T \vartheta(s)dW(s) - \frac{1}{2}\int_0^T \vartheta^2(s)ds\right) + w + \frac{1}{\beta}H(\mathbf{P}^*|\mathbf{P}), \tag{14.20}$$

where $H(\mathbf{P}^*|\mathbf{P}) = \mathbf{E}\left(\varphi(T)\log\varphi(T)\right)$, the condition (14.19) supplies the existence of $H(\mathbf{P}^*|\mathbf{P})$, and the maximal value of the expected utility is

$$\mathbf{E}(u(X^*)) = 1 - \exp\left\{-\beta w - H(\mathbf{P}^*|\mathbf{P})\right\}.$$

Let $\varphi(T)$ be chosen in such a way that the corresponding process ϑ satisfies the assumption of Lemma 14.2. Also, let W-transformable process G satisfy conditions (A) and (B) of Theorem 14.4, and ϑ satisfy (14.13) with $p > 1$. Then we can conclude directly from representation (14.20) that conditions of Theorem 14.4 hold. Therefore, the optimal profile X^* admits the representation $X^* = \int_0^T \psi(s)dG(s)$.

Remark 14.5 Similarly, under the same conditions as above, we can conclude that for any constant $d \in \mathbb{R}$ there exists ψ_d such that $X^* = d + \int_0^T \psi_d(s)dG(s)$. Therefore, we can start from any initial value of the capital and achieve the desirable wealth. In this sense, w is not necessarily the initial wealth as it is often assumed in the semimartingale framework, but is rather a budget constraint in the generalized sense.

Remark 14.6 In the case when W-transformable Gaussian process G is a semimartingale, we can use Girsanov's theorem in order to get the representation, similar to (14.17). Indeed, let, for example, G be a Gaussian process of the form $G(t) = \int_0^t \mu(s)ds + \int_0^t a(s)dW(s)$, $|\mu(s)| \le \mu$, $a(s) > a > 0$ are non-random measurable functions, and ξ is $\mathcal{F}_T^W$-measurable random variable, $\mathbf{E}(\xi^2) < \infty$. Then we transform G into $\widetilde{G} = \int_0^{\cdot} a(s)d\widetilde{W}(s)$, with the help of equivalent probability measure $\widetilde{\mathbf{P}}$ having Radon–Nikodym derivative

$$\frac{d\widetilde{\mathbf{P}}}{d\mathbf{P}} = \exp\left\{-\int_0^T \frac{\mu(s)}{a(s)} dW(s) - \frac{1}{2}\int_0^T \left(\frac{\mu(s)}{a(s)}\right)^2 ds\right\}.$$

With respect to this measure $\mathbf{E}_{\widetilde{\mathbf{P}}}|X^*| < \infty$, and we get the following representation

$$\begin{aligned} X^* =& \mathbf{E}_{\widetilde{\mathbf{P}}}(X^*) + \int_0^T \psi(s) d\widetilde{W}_s = \mathbf{E}_{\widetilde{\mathbf{P}}}(X^*) + \int_0^T \frac{\psi(s)}{a(s)} d\widetilde{G}(s) \\ =& \mathbf{E}_{\widetilde{\mathbf{P}}}(X^*) + \int_0^T \frac{\psi(s)}{a(s)} dG(s) = \mathbf{E}_{\widetilde{\mathbf{P}}}(X^*) + \int_0^T \psi(s)\mu(s) ds + \int_0^T \psi(s) dW(s). \end{aligned} \tag{14.21}$$

Representations (14.17) and (14.21) have the following distinction: (14.17) "starts" from 0 (but can start from any other constant) while (14.21) "starts" exactly from $\mathbf{E}_{\widetilde{\mathbf{P}}}(X^*)$.

As we can see, the solution of the utility maximization problem for W–transformable process depends on the process in indirect way, through the random variable $\varphi(T)$ such that $\mathbf{E}\varphi(T) = 1$, $\varphi(T) > 0$ a.s. Also, this solution depends on whether or not we can choose the appropriate value of c, but this is more or less a technical issue. Let us return to the choice of $\varphi(T)$. In the case of the semimartingale market, $\varphi(T)$ can be reasonably chosen as the likelihood ratio of some martingale measure, and the choice is unique in the case of the complete market. The non-semimartingale market can contain some hidden semimartingale structure. To illustrate this, consider two examples.

Example 14.2 Let the market consist of bond B and stock S,

$$B(t) = e^{rt}, \ S(t) = \exp\left\{\mu t + \sigma B_t^H\right\},$$

$r \geq 0$, $\mu \in \mathbb{R}$, $\sigma > 0$, $H > \frac{1}{2}$. The discounted price process has a form $Y(t) = \exp\left\{(\mu - r)t + \sigma B_t^H\right\}$. It is well-known that such market admits an arbitrage, but even in these circumstances the utility maximization problem makes sense. Well, how to choose $\varphi(T)$? There are at least two natural approaches.

1. Note that for $H > \frac{1}{2}$ the kernel K^H from (14.14) has a form

$$K^H(t,s) = C(H) s^{\frac{1}{2}-H} \int_s^t u^{H-\frac{1}{2}} (u-s)^{H-\frac{3}{2}} du,$$

and representation (14.15) has a form

$$W(t) = (C(H))^{-1} \int_0^t s^{\frac{1}{2}-H} K^*(t,s) dB_s^H,$$

where

$$K^*(t,s) = \left(t^{H-\frac{1}{2}}(t-s)^{\frac{1}{2}-H} - \left(H-\frac{1}{2}\right)\int_s^t u^{H-\frac{3}{2}}(u-s)^{\frac{1}{2}-H}du\right)\frac{1}{\Gamma\left(\frac{3}{2}-H\right)}.$$

Therefore,

$$\begin{aligned}
&(C(H))^{-1}\int_0^t s^{\frac{1}{2}-H}K^*(t,s)d\left((\mu-r)s+\sigma B_s^H\right)\\
&=\sigma W(t)+\frac{\mu-r}{C(H)}\int_0^t s^{\frac{1}{2}-H}K^*(t,s)ds\\
&=\sigma W(t)+\frac{\mu-r}{C(H)\Gamma\left(\frac{3}{2}-H\right)}\int_0^t\left(s^{\frac{1}{2}-H}t^{H-\frac{1}{2}}(t-s)^{\frac{1}{2}-H}\right.\\
&\quad\left.-\left(H-\frac{1}{2}\right)s^{\frac{1}{2}-H}\int_s^t u^{H-\frac{3}{2}}(u-s)^{\frac{1}{2}-H}du\right)ds\\
&=\sigma W_t+\frac{\mu-r}{C(H)\Gamma(\frac{3}{2}-H)}\frac{\Gamma^2(\frac{3}{2}-H)}{(\frac{3}{2}-H)\Gamma(2-2H)}t^{\frac{3}{2}-H}\\
&=\sigma W_t+(\mu-r)C_1(H)t^{\frac{3}{2}-H},
\end{aligned}$$

where

$$C_1(H)=\left(\frac{3}{2}-H\right)^{-1}\left(\frac{\Gamma(\frac{3}{2}-H)}{2H\Gamma(2-2H)\Gamma(H+\frac{1}{2})}\right)^{\frac{1}{2}}.$$

In this sense we say that the model involves a hidden semimartingale structure.

Consider a virtual semimartingale asset

$$\begin{aligned}
\hat{Y}(t)&=\exp\left\{(C(H))^{-1}\int_0^t s^{\frac{1}{2}-H}K^*(t,s)d\log Y(s)\right\}\\
&=\exp\left\{\sigma W_t+(\mu-r)C(H)t^{\frac{3}{2}-H}\right\}.
\end{aligned}$$

We see that measure $\mathbf{P}^*$ such that

$$\begin{aligned}
\frac{d\mathbf{P}^*}{d\mathbf{P}}=\exp\left\{-\int_0^T\left(\frac{(\mu-r)C_2(H)}{\sigma}s^{\frac{1}{2}-H}+\frac{\sigma}{2}\right)dW_s\right.\\
\left.-\frac{1}{2}\int_0^T\left(\frac{(\mu-r)C_2(H)}{\sigma}s^{\frac{1}{2}-H}+\frac{\sigma}{2}\right)^2ds\right\},
\end{aligned}\tag{14.22}$$

where $C_2(H)=C_1(H)\left(\frac{3}{2}-H\right)$, reduces $\hat{Y}(t)$ to the martingale of the form $\exp\left\{\sigma W_t-\frac{\sigma^2}{2}t\right\}$. Therefore, we can put $\varphi(T)=\frac{d\mathbf{P}^*}{d\mathbf{P}}$ from (14.22). Regarding the

Hölder property, $\vartheta(s) = s^{\frac{1}{2}-H}$ satisfies (14.13) with some $p > 1$ for any $H \in (\frac{1}{2}, 1)$. Therefore, for the utility function $u(x) = 1 - e^{-\alpha x}$ we have

$$X^* = \frac{1}{\alpha}\left(\int_0^T \varsigma(s)dW_s - \frac{1}{2}\int_0^T \varsigma_s^2 ds\right) + W + \frac{1}{2}H(\mathbf{P}^*|\mathbf{P}),$$

where $\varsigma(s) = \frac{(\mu-r)C_2(H)}{\sigma}s^{\frac{1}{2}-H} + \frac{\sigma}{2}$, and $|H(\mathbf{P}^*|\mathbf{P})| < \infty$.

2. It was proved in [8] that the fractional Brownian motion B^H is the limit in $L_p(\Omega, \mathcal{F}, \mathbf{P})$ for any $p > 0$ of the process

$$B^{H,\varepsilon}(t) = \int_0^t K(s+\varepsilon, s)dW(s) + \int_0^t \psi_\varepsilon(s)ds,$$

where W is he underlying Wiener process, i.e. $B^H(t) = \int_0^t K(t,s)dW(s)$, where

$$K(t,s) = C_H s^{\frac{1}{2}-H}\int_s^t u^{H-\frac{1}{2}}(u-s)^{H-\frac{3}{2}}du,$$

$$\psi_\varepsilon(s) = \int_0^s \partial_1 K(s+\varepsilon, u)dW_u,$$

$$\partial_1 K(t,s) = \frac{\partial K(t,s)}{\partial t} = C_H s^{\frac{1}{2}-H}t^{H-\frac{1}{2}}(t-s)^{H-\frac{3}{2}}.$$

Consider prelimit market with discounted risky asset price Y^ε of the form

$$Y^\varepsilon(t) = \exp\left\{(\mu - r)t + \sigma\int_0^t \psi_\varepsilon(s)ds + \sigma\int_0^t K(s+\varepsilon, s)dW_s\right\}.$$

This financial market is arbitrage-free and complete, and the unique martingale measure has the Radon-Nikodym derivative

$$\varphi_\varepsilon(T) = \exp\left\{-\int_0^T \zeta_\varepsilon(t)dW_t - \frac{1}{2}\int_0^T \zeta_\varepsilon^2(t)dt\right\},$$

where

$$\zeta_\varepsilon(t) = \frac{\mu - r + \sigma\psi_\varepsilon(t)}{\sigma K(t+\varepsilon, t)} + \frac{1}{2}\sigma K(t+\varepsilon, t).$$

Note that $K(t+\varepsilon, t) \to 0$ as $\varepsilon \to 0$. Furthermore, $\rho_t = \frac{\mu-r+\sigma\psi_\varepsilon(t)}{\sigma K(t+\varepsilon,t)}$ is a Gaussian process with $\mathbf{E}\rho_t = 0$ and

$$\mathbf{var}\zeta_\varepsilon(t) = \int_0^t \left(\frac{\partial_1 K(t+\varepsilon,u)}{K(t+\varepsilon,t)}\right)^2 du$$
$$= \int_0^t \left(\frac{u^{1/2-H}(t+\varepsilon)^{H-1/2}(t+\varepsilon-u)^{H-3/2}}{t^{1/2-H}\int_t^{t+\varepsilon} v^{H-1/2}(v-t)^{H-3/2}}\right)^2 du$$
$$\geq \varepsilon^{1-2H}\int_0^t (t+\varepsilon-u)^{2H-3} du = \frac{\varepsilon^{1-2H}t}{2-2H}\left(\varepsilon^{2H-2} - (t+\varepsilon)^{2H-2}\right) \to \infty.$$

Therefore, we can not get a reasonable limit of $\varphi_\varepsilon(T)$ as $\varepsilon \to 0$. Thus one should use this approach with great caution.

14.4.2 *Expected Utility Maximization for Restricted Capital Profiles*

Consider now the case when the utility function u is defined on some interval (a, ∞). Assume for technical simplicity that $a = 0$. Therefore, in this case the set $\mathcal{B}_0$ of admissible capital profiles has a form

$$\mathcal{B}_0 = \left\{X \in L^0(\Omega, \mathcal{F}, \mathbf{P}) : X \geq 0 \text{ a.s. and } \mathbf{E}(\varphi(T)X) = w\right\}.$$

Assume that the utility function u is continuously differentiable on $(0, \infty)$, introduce $\pi_1 = \lim\limits_{x\uparrow\infty} u'(x) \geq 0, \pi_2 = u'(0+) = \lim\limits_{x\downarrow 0} u'(x) \leq +\infty$, and define $I^+ : (\pi_1, \pi_2) \longrightarrow (0, \infty)$ as the continuous, bijective function, inverse to u' on (π_1, π_2).

Extend I^+ to the whole half-axis $[0, \infty]$ by setting

$$I^+(y) = \begin{cases} +\infty, & y \leq \pi_1 \\ 0, & y \geq \pi_2. \end{cases}$$

Theorem 14.6 ([10], Theorem 3.39) *Let the random variable $X^* \in \mathcal{B}_0$ have a form $X^* = I^+(c\varphi(T))$ for a constant $c > 0$ such that*

$$\mathbf{E}(\varphi(T)I^+(c\varphi(T))) = w.$$

If $\mathbf{E}u(X^*) < \infty$, then

$$\mathbf{E}(u(X^*)) = \max_{X\in\mathcal{B}_0} \mathbf{E}(u(X)),$$

and this maximizer is unique.

From here we deduce the corresponding result on the solution of utility maximization problem similarly to Theorem 14.5. Define, as before,

$$\mathcal{B}_w^G = \left\{ \psi : [0,T] \times \Omega \to \mathbb{R} \;\middle|\; \psi \text{ is bounded } \mathcal{F}_t^W\text{-adapted, there exists a generalized Lebesgue-Stieltjes integral } \int_0^T \psi(s)dG(s) \geq 0, \text{ and } \mathbf{E}\left(\varphi(T)\int_0^T \psi(s)dG(s)\right) = w \right\}.$$

Theorem 14.7 *Let the following conditions hold*

(i) Gaussian process G satisfies conditions (A) and (B).
(ii) Function $I^+(x), x \in \mathbb{R}$ is Hölder continuous.
(iii) Stochastic process ϑ in representation (14.16) satisfies (14.13) with some $p > 1$.
(iv) There exists $c \in \mathbb{R}$ such that $\mathbf{E}(\varphi(T)I^+(c\varphi(T))) = w$.

Then the random variable $X^* = I^+(c\varphi(T))$ admits the representation

$$X^* = \int_0^T \overline{\psi}(s)dG(s)$$

with some $\overline{\psi} \in \widetilde{\mathcal{B}}_w^G$. If $\mathbf{E}u(X^*) < \infty$, X^* is the solution to expected utility maximization problem:

$$\mathbf{E}(u(X^*)) = \max_{\psi \in \widetilde{\mathcal{B}}_w^G} \mathbf{E}\left(u\left(\int_0^T \psi(s)dG(s)\right)\right).$$

Example 14.3 Consider the case of CARA utility function u. Let first $u(x) = \frac{x^\gamma}{\gamma}$, $x > 0, \gamma \in (0,1)$. Then, according to [10, Example 3.43],

$$I^+(c\varphi(T)) = c^{-\frac{1}{1-\gamma}}(\varphi(T))^{-\frac{1}{1-\gamma}}.$$

If $d := \mathbf{E}(\varphi(T))^{-\frac{\gamma}{1-\gamma}} < \infty$, then the unique optimal profile is given by $X^* = \frac{w}{d}(\varphi(T))^{-\frac{1}{1-\gamma}}$, and the maximal value of the expected utility is equal to

$$\mathbf{E}(u(X^*)) = \frac{1}{\gamma} w^\gamma d^{1-\gamma}.$$

As it was mentioned,

$$\varphi = \varphi(T) = \exp\left\{\int_0^T \vartheta(s)dW(s) - \frac{1}{2}\int_0^T \vartheta^2(s)ds\right\}, \tag{14.23}$$

thus

$$(\varphi(T))^{-\frac{1}{1-\gamma}} = \exp\left\{-\frac{1}{1-\gamma}\int_0^T \vartheta(s)dW(s) + \frac{1}{2(1-\gamma)}\int_0^T \vartheta^2(s)ds\right\}.$$

Therefore, we get the following result.

Theorem 14.8 *Let the process ϑ in the representation* (14.23) *satisfy* (14.13)*, and*

$$\mathbf{E}\exp\left\{-\frac{\gamma}{1-\gamma}\int_0^T \vartheta(s)dW_s + \frac{\gamma}{2(1-\gamma)}\int_0^T \vartheta_s^2 ds\right\} < \infty.$$

Let the process G satisfy the same conditions as in Theorem 14.5. Then $X^* = \int_0^T \psi(s)dG(s)$.

In the case where $u(x) = \log x$, we have $\gamma = 0$ and $X^* = \frac{w}{\varphi(T)}$. Assuming that the relative entropy $H(\mathbf{P}|\mathbf{P}^*) = \mathbf{E}(\frac{1}{\varphi(T)}\log\varphi(T))$ is finite, we get that

$$\mathbf{E}(\log X^*) = \log w + H\left(\mathbf{P}|\mathbf{P}^*\right).$$

14.5 Conclusion

We have studied a broad class of non-semimartingale financial market models, where the random drivers are Wiener-transformable Gaussian random processes, i.e. some adapted transformations of a Wiener process. Under assumptions that the incremental variance of the process satisfies two-sided power bounds, we have given sufficient conditions for random variables to admit integral representations with bounded adapted integrand; these representations are models for bounded replicating strategies. It turned out that these representation results can be applied to solve utility maximization problems in non-semimartingale market models.

Acknowledgements Elena Boguslavskaya is supported by Daphne Jackson fellowship funded by EPSRC. The research of Yu. Mishura was funded (partially) by the Australian Government through the Australian Research Council (project number DP150102758). Yu. Mishura acknowledges that the present research is carried through within the frame and support of the ToppForsk project nr. 274410 of the Research Council of Norway with title STORM: Stochastics for Time-Space Risk Models.

References

1. Androshchuk, T., Mishura, Y.: Mixed Brownian-fractional Brownian model: absence of arbitrage and related topics. Stochastics **78**(5), 281–300 (2006)
2. Bender, C., Sottinen, T., Valkeila, E.: Pricing by hedging and no-arbitrage beyond semimartingales. Financ. Stoch. **12**(4), 441–468 (2008)
3. Bender, C., Sottinen, T., Valkeila, E.: Fractional processes as models in stochastic finance. In: Advanced Mathematical Methods for Finance, pp. 75–103. Springer, Heidelberg (2011)
4. Björk, T.: Arbitrage Theory in Continuous Time, 2nd edn. Oxford University Press, Oxford (2004)
5. Cheridito, P.: Mixed fractional Brownian motion. Bernoulli **7**(6), 913–934 (2001)
6. Cheridito, P.: Arbitrage in fractional Brownian motion models. Financ. Stoch. **7**(4), 533–553 (2003)
7. Dudley, R.M.: Wiener functionals as Itô integrals. Ann. Probab. **5**(1), 140–141 (1977)
8. Dung, N.T.: Semimartingale approximation of fractional Brownian motion and its applications. Comput. Math. Appl. **61**(7), 1844–1854 (2011)
9. Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. Translated from the French, Studies in Mathematics and its Applications, vol. 1. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York (1976)
10. Föllmer, H., Schied, A.: Stochastic Finance. An Introduction in Discrete Time, Extended edn. Walter de Gruyter & Co., Berlin (2011)
11. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics, vol. 113, 2nd edn. Springer, New York (1991)
12. Karatzas, I., Shreve, S.E.: Methods of Mathematical Finance, Applications of Mathematics (New York), vol. 39. Springer, New York (1998)
13. Li, W.V., Shao, Q.M.: Gaussian processes: inequalities, small ball probabilities and applications. In: Handbook of Statistics, vol. 19, pp. 533–597 (2001)
14. Lifshits, M.A.: Gaussian Random Functions, Mathematics and its Applications, vol. 322. Kluwer Academic Publishers, Dordrecht (1995)
15. Mishura, Y., Shevchenko, G.: Small ball properties and representation results. Stoch. Process. Appl. **127**(1), 20–36 (2017)
16. Mishura, Y., Shevchenko, G., Valkeila, E.: Random variables as pathwise integrals with respect to fractional Brownian motion. Stoch. Process. Appl. **123**(6), 2353–2369 (2013)
17. Mishura, Y.S.: Stochastic Calculus for Fractional Brownian Motion and Related Processes. Lecture Notes in Mathematics, vol. 1929. Springer, Berlin (2008)
18. Norros, I., Valkeila, E., Virtamo, J.: An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. Bernoulli **5**(4), 571–587 (1999)
19. Rogers, L.C.G.: Arbitrage with fractional Brownian motion. Math. Financ. **7**(1), 95–105 (1997)
20. Samko, S.G., Ross, B.: Integration and differentiation to a variable fractional order. Integral Transform. Spec. Funct. **1**(4), 277–300 (1993)
21. Shalaiko, T., Shevchenko, G.: Integral representation with respect to fractional brownian motion under a log-hölder assumption. Modern Stoch.: Theory Appl. **2**(3), 219–232 (2015)
22. Shevchenko, G., Viitasaari, L.: Adapted integral representations of random variables. Int. J. Modern Phys.: Conf. Ser. **36**, Article ID 1560004, 16 (2015)
23. Zähle, M.: On the link between fractional and stochastic calculus. In: Stochastic Dynamics (Bremen, 1997), pp. 305–325. Springer, New York (1999)

Chapter 15
A New Approach to the Modeling of Financial Volumes

Guglielmo D'Amico, Fulvio Gismondi and Filippo Petroni

Abstract In this paper we study the high frequency dynamic of financial volumes of traded stocks by using a semi-Markov approach. More precisely we assume that the intraday logarithmic change of volume is described by a weighted-indexed semi-Markov chain model. Based on this assumptions we show that this model is able to reproduce several empirical facts about volume evolution like time series dependence, intra-daily periodicity and volume asymmetry. Results have been obtained from a real data application to high frequency data from the Italian stock market from first of January 2007 until end of December 2010.

Keywords Semi-Markov process · High frequency data · Financial volume

15.1 Introduction

Studies on market microstructure have acquired a crucial importance in order to explain the price formation process, see e.g. De Jong and Rindi [8]. The main variables are (logarithmic) price returns, volumes and duration. Sometimes they are modeled jointly and the main approach is the so-called econometric analysis, see e.g. Manganelli [14] and Podobnik et al. [16] and the bibliography therein.

G. D'Amico (✉)
Department of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, via dei vestini 31, 66013 Chieti, Italy
e-mail: g.damico@unich.it

F. Gismondi
Department of Economic and Business Science, University "Guglielmo Marconi", Via Plinio 44, 00193 Rome, Italy
e-mail: f.gismondi@unimarconi.it

F. Petroni
Department of Economy and Business, University of Cagliari, via SantâĂŹIgnazio, 17, 09123 Cagliari, Italy
e-mail: fpetroni@unica.it

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*, Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_15

The volume variable is very important not only because it interacts directly with duration and returns but also because a correct specification of forecasted volumes can be used for Volume Weighted Average Price trading, see e.g. Brownlees et al. [1]. This variable has been investigated for long time and several statistical regularities have been highlighted, see e.g. Jain and Joh [11], Gopikrishnan et al. [10], Lobato and Velasco [12] and more recently Plerou and Stanley [15].

In this paper we propose an alternative approach to the modeling of financial volumes which is based on a generalization of semi-Markov processes called Weighted-Indexed Semi-Markov Chain (WISMC) model. This choice is motivated by recent results in the modeling of price returns in high-frequency financial data where the WISMC approach was demonstrated to be particular efficient in reproducing the statistical properties of financial returns, see D'Amico and Petroni [2–6]. The WISMC model is very flexible and for this reason we decided to test its appropriateness also for financial volumes.

It should be remarked that WISMC models generalize semi-Markov processes and also non-Markovian models based on continuous time random walks that were used extensively in the econophysics community, see e.g. Mainardi et al. [13] and Raberto et al. [17] as well as in actuarial sciences, see e.g. Stenberg et al. [19], D'Amico et al. [7], and Silvestrov et al. [18].

The model is applied to a database of high frequency volume data from all the stocks in the Italian Stock Market from first of January 2007 until end of December 2010. In the empirical analysis we find that WISMC model is a good choice for modeling financial volumes which is able to correctly reproduce stylized facts documented in literature of volumes such as the autocorrelation function.

The paper is divided as follows: First, Weighted-Indexed-Semi-Markov chains are shortly described in Sect. 15.2. Next, we introduce our model of financial volumes and an application to real high frequency data illustrates the results. Section 15.4 concludes and suggests new directions for future developments.

15.2 Weighted-Indexed Semi-Markov Chains

The general formulation of the WISMC as developed in D'Amico and Petroni [3–5] is here only discussed informally.

WISMC models share similar ideas as those that generate Markov processes and semi-Markov processes. These processes are all described by a set of finite states E and a sequence of random variables $\{J_n\}_{n\in\mathbb{N}}$ denoting the successive states visited by the system whose transitions are ruled by a transition probability matrix. The semi-Markov process differs from the Markov process because the transition times T_n where the states change values, are generated according to random variables. Indeed, the time between transitions $T_{n+1} - T_n$ is random and may be modeled by means of any type of distribution functions. In order to better represent the statistical characteristics of high-frequency financial data, in a recent article, the idea of a WISMC model was introduced in the field of price returns, see D'Amico and Petroni

[3, 4]. The novelty, with respect to the semi-Markov case, consists in the introduction of a third random variable defined as follows:

$$I_n(\lambda) = \sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} f^{\lambda}(J_{n-1-k}, T_n, a), \tag{15.1}$$

where f is any real value bounded function and I_0^{λ} is known and non-random. The variable I_n^{λ} is designated to summarize the information contained in the past trajectory of the $\{J_n\}$ process that is relevant for future predictions. Indeed, at each past state J_{n-1-k} occurred at time $a \in \mathbb{N}$ is associated the value $f^{\lambda}(J_{n-1-k}, T_n, a)$, which depends also on the current time T_n. The quantity λ denotes a parameter that represents a weight and should be calibrated to data. In the applicative section we will describe the calibration of λ as well as the choice of the function f.

The WISMC model is specified once a dependence structure between the variables is considered. Toward this end, the following assumption is formulated:

$$\begin{aligned} &\mathbb{P}[J_{n+1} = j,\ T_{n+1} - T_n \le t | J_n, T_n, I_n^{\lambda}, J_{n-1}, T_{n-1}, I_{n-1}^{\lambda}, \ldots] \\ &\quad = \mathbb{P}[J_{n+1} = j,\ T_{n+1} - T_n \le t | J_n, I_n^{\lambda}] := Q_{J_n j}^{\lambda}(I_n^{\lambda}; t). \end{aligned} \tag{15.2}$$

Relation (15.2) asserts that the knowledge of the values of the variables J_n, I_n^{λ} is sufficient to give the conditional distribution of the couple $J_{n+1}, T_{n+1} - T_n$ whatever the values of the past variables might be. Therefore to make probabilistic forecasting we need the knowledge of the last state of the system and the last value of the index process. If $\mathbf{Q}^{\lambda}(x; t)$ is constant in x then the WISMC kernel degenerates in an ordinary semi-Markov kernel and the WISMC model becomes equivalent to classical semi-Markov chain models, see e.g. D'Amico and Petroni [3] and Fodra and Pham [9].

The probabilities $Q_{ij}^{\lambda}(x; t))_{i,j \in E}$ can be estimated directly using real data. In D'Amico and Petroni [6] it is shown that the estimator $\hat{Q}_{i,j}^{\lambda}(x; t) := \frac{N_{ij}(x;t)}{N_i(x)}$, is the approached maximum likelihood estimator of the corresponding transition probabilities. The quantity $N_{ij}(x; t)$ expresses the number of transitions from state i, with an index value x, to state j with a sojourn time in state i less or equal to t. The quantity $N_i(x)$ is the number of visits to state i with an index value x.

Once the WISMC kernel $\mathbf{Q}^{\lambda}(x; t)$ is known, it is possible to compute the state probabilities of the system at any time $t \in \mathbb{N}$. To show how to compute transition probability functions we define by

$$\begin{aligned} &N(t) = \sup\{n \in \mathbb{N} : T_n \le t\};\quad Z(t) = J_{N(t)}; \\ &I(\lambda; t) = \sum_{k=0}^{N(t)-1+\theta} \sum_{a=T_{N(t)+\theta-1-k}}^{(t \wedge T_{N(t)+\theta-k})-1} f^{\lambda}(J_{N(t)+\theta-1-k}, t, a), \end{aligned} \tag{15.3}$$

where $\theta = 1_{\{t > T_{N(t)}\}}$.

The stochastic processes defined in (15.3) represent the number of transitions up to time t, the state of the system at time t and the value of the index process up to t, respectively. We refer to $Z(t)$ as a weighted indexed semi-Markov process.

As it is well known, it is possible to give an alternative description of the semi-Markov process by introducing the backward recurrence time process $B(t) := t - T_{N(t)}$ and to describe the probabilistic behavior of the Markov process $(Z(t), B(t))$ on the extended state space $E \times \overline{\mathbb{N}}$ where $\overline{\mathbb{N}} = \{0, 1, ..., N\}$ and N is the maximum length of stay of the states of the process. This technique was first proposed in Vassiliou and Papadopoulou [20] and proved useful is studying certain aspects of non-homogeneous semi-Markov process. Also in our more general setting it is possible to describe the system behavior by using the backward recurrence time process, this choice is adopted here to have a description of the one-step transition probabilities of the WISMC model and result to be very useful in the next section for the definition of the bivariate model. Let denote by

$$p_{((i,u)(j,d))}(v) := \mathbb{P}[Z(n+1) = j,\ B(n+1) = d \mid Z(n) = i,\ B(n) = u,\ I(\lambda; n) = v]. \tag{15.4}$$

It should be noted that $d \in \{0, u+1\}$ because if a transition from state i to an arbitrary state j, $j \neq i$ is executed, then $B(n+1) = n + 1 - T_{N(n+1)} = n + 1 - (n+1) = 0$. On the contrary, if the system will stay in state i next period, then, being $T_{N(n+1)} = T_{N(n)}$, we have $B(n+1) = n + 1 - T_{N(n+1)} = 1 + (n - T_{N(n)}) = 1 - B(n) = 1 + u$.

The probabilities (15.4) can be obtained from the indexed semi-Markov kernel, to proove this, we first need to give the following

Lemma 15.1 (see D'Amico and Petroni [5]) *Let suppose that $I(\lambda; n) = v$, $T_{N(n)} = n - u$ with $n \geq u \geq 0$ and $T_{N(n)+1} > n$, then*

$$\begin{aligned} I_{N(n)}(\lambda) = v - \sum_{a=T_{N(n)}}^{n-1} f^{\lambda}(J_{N(n)}, n, a) \\ + \sum_{k=0}^{N(n)-1} \sum_{a=T_{N(n)-1-k}}^{T_{N(n)-k}-1} \Delta f^{\lambda}(J_{N(n)-k}, T_{N(n)}, n, a), \end{aligned} \tag{15.5}$$

where $\Delta f^{\lambda}(i, T_{N(n)}, n, a) := f^{\lambda}(i, T_{N(n)}, a) - f^{\lambda}(i, n, a)$.

Theorem 15.1 (see D'Amico and Petroni [5]) *For all $i, j \in E, u, d \in \mathbb{N}$ and $v \in R$, the one step transition probabilities $p_{((i,u)(j,d))}(v)$ are given by*

$$p_{((i,u)(j,d))}(v) = \begin{cases} \frac{\bar{H}_i^{\lambda}(v+\Delta I(N(n),n);1+u)}{\bar{H}_i^{\lambda}(v+\Delta I(N(n),n);u)} & \text{if } j = i,\ d = 1+u \\ \frac{q_{ij}^{\lambda}(v+\Delta I(N(n),n);1+u)}{\bar{H}_i^{\lambda}(v+\Delta I(N(n),n);u)} & \text{if } j \neq i,\ d = 0, \\ 0 & \text{otherwise}, \end{cases} \tag{15.6}$$

where $\bar{H}_i^{\lambda}(v;t) = 1 - H_i^{\lambda}(v;t)$ *is the survival function of sojourn time in state* i, $q_{ij}^{\lambda}(x,t) = Q_{ij}^{\lambda}(x,t) - Q_{ij}^{\lambda}(x,t-1)$ *and* $\Delta I(N(n),n) = I_{N(n)}(\lambda) - I_n(\lambda)$ *is the opposit of the increment of the index process on the waiting time* $n - N(n)$.

It should be remarked that the computation of the probabilities (15.4) can be done through formula (15.6) where it is necessary to evaluate the quantity $\Delta I(N(n),n)$. This last quantity is obtained thanks to Lemma 15.1 and has to be recalculated step by step.

15.3 The Volume Model

Let us assume that the trading volume of the asset under study is described by the time varying process $V(t)$, $t \in \mathbb{N}$.

The (logarithmic) change of volume at time t over the unitary time interval is defined by

$$Z^V(t) = \log \frac{V(t+1)}{V(t)}. \tag{15.7}$$

On a short time scale, $Z^V(t)$ changes value in correspondence of an increasing sequence of random times, $\{T_n^V\}_{n\in\mathbb{N}}$. According to the notation adopted in the previous section, we denote the values assumed at time T_n^V by J_n^V and the corresponding values of the index process by

$$I_n^V(\lambda) = \sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}^V}^{T_{n-k}^V - 1} f^{\lambda}(J_{n-1-k}^V, T_n^V, a). \tag{15.8}$$

If we assume that the variables $(J_n^V, T_n^V, I_n^V(\lambda))$ satisfy relationship (15.2) then the volume process can be described by a WISMC model and if $N^V(t) = \sup\{n \in \mathbb{N} : T_n^V \le t\}$ is the number of transition of the volume process, then $Z^V(t) = J_{N^V(t)}^V$ is the WISMC process that describes the volume values at any time t.

We assume also that the set E is finite and is obtained by an opportune discretization of the values of the financial volumes. A description of the adopted discretization and of the state space model is described in next section. The choice of a finite state space can be formalized by defining the state space by $E = \{-z_{min}\Delta, \ldots, -2\Delta, -\Delta, 0, \Delta, 2\Delta, \ldots, z_{max}\Delta\}$.

Our objective is to demonstrate that a WISMC model for financial volume is able to reproduce some important facts of volume dynamics. One of the most important feature is the persistence of the volume process. As found in Manganelli [14], the volume process is strongly persistent for frequently traded stocks. A possible explanations is that the volume is pushed up by the arrival of new information to the market participants and some times is necessary in order to delete the effects of the information arrival. In order to investigate the goodness of the WISMC model we

define the autocorrelation of the modulus of volumes as

$$\Sigma(t, t+\tau) = Cov(\mid Z^V(t+\tau) \mid, \mid Z^V(t) \mid). \tag{15.9}$$

Another important statistic is the first passage time distribution of the volume process. First we need to define the accumulation factor of the volume process from the generic time t to time $t+\tau$: $M_t^V(\tau) = \frac{V(t+\tau)}{V(t)} = e^{\sum_{r=0}^{\tau-1} Z^V(t+r)}$. We will denote the fpt by

$$\Gamma_\sigma = \min\{\tau \geq 0; M_0^V(\tau) \geq \sigma\}, \tag{15.10}$$

where σ is a given threshold. We are interested in finding the distributional properties of the fpt, that is to compute $\mathbb{P}[\Gamma_\sigma > t | (J^V, T^V, I^V(\lambda))_{-m}^0 = (i,t)_{-m}^0]$, where $(\mathbf{J^V})_{-m}^0 = (J_{-m}^V, J_{-m+1}^V, \ldots, J_0^V)$, $(\mathbf{T^V})_{-m}^0 = (T_{-m}^V, T_{-m+1}^V, \ldots, T_0^V)$,

$$(\mathbf{I^V}(\lambda))_{-m}^0 = (I_{-m}^V(\lambda), I_{-m+1}^V(\lambda), \ldots, I_0^V(\lambda)).$$

15.4 Application to Real High Frequency Data

The data used in this work are tick-by-tick quotes of indexes and stocks downloaded from www.borsaitaliana.it for the period January 2007–December 2010 (4 full years). The data have been re-sampled to have 1 min frequency. Every minutes the last price and the cumulated volume (number of transaction) is recorded. For each stock the database is composed of about $5 * 10^5$ volumes and prices. The list of stocks analyzed and their symbols are reported in Table 15.1.

In Figs. 15.1 and 15.2 are shown the trading volume and the logarithmic change of the trading volume for the 4 stocks in the analysed period. From both figures it is possible to notice that there are period of high volume, essentially when the trading frequency is higher, followed by periods of low volumes showing a certain degree of clustering and autocorrelation.

To model $Z^V(t)$ as a WISMC model the first step is to discretize the variable. We choose 5 states of discretization with the following edges: $\{-\infty, -4, -1, 1, 4, \infty\}$. The number of times that $Z^V(t)$ fall in each state is shown in Fig. 15.3. For all stocks the frequency of each state ie symmetric with respect to state 3.

Table 15.1 Stocks used in the application and their symbols

F	Fiat
ISP	Intesa San Paolo
TIT	Telecom
TEN	Tenaris

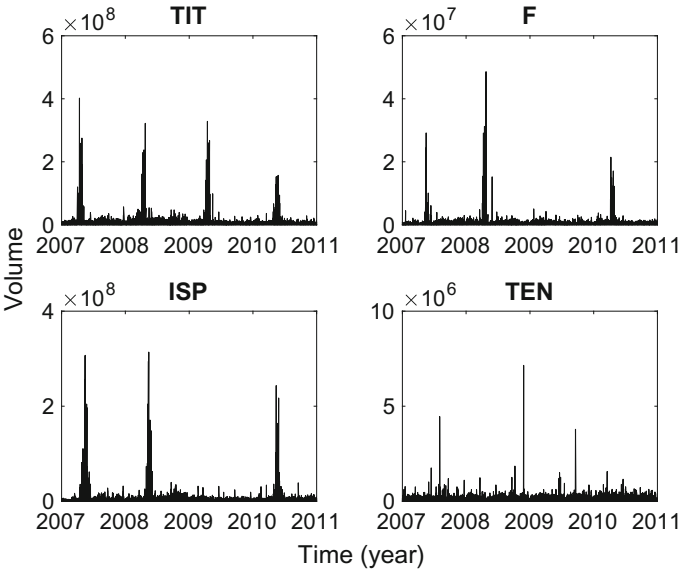

Fig. 15.1 Trading volume of the analyzed stocks

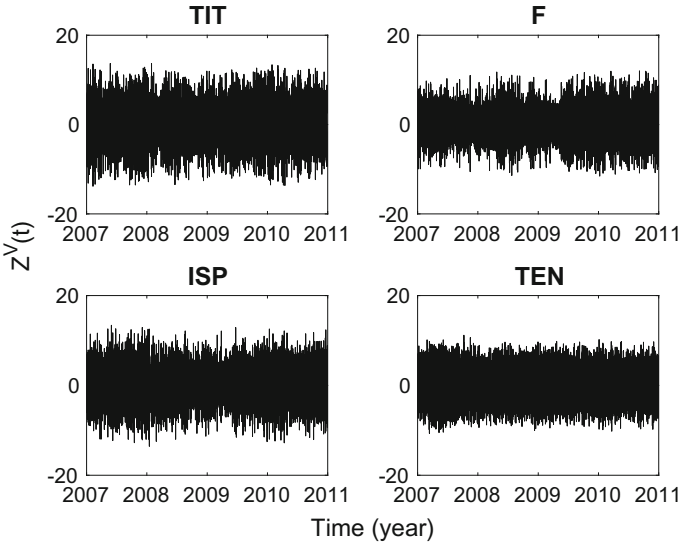

Fig. 15.2 The (logarithmic) change of trading volume of the analyzed stocks

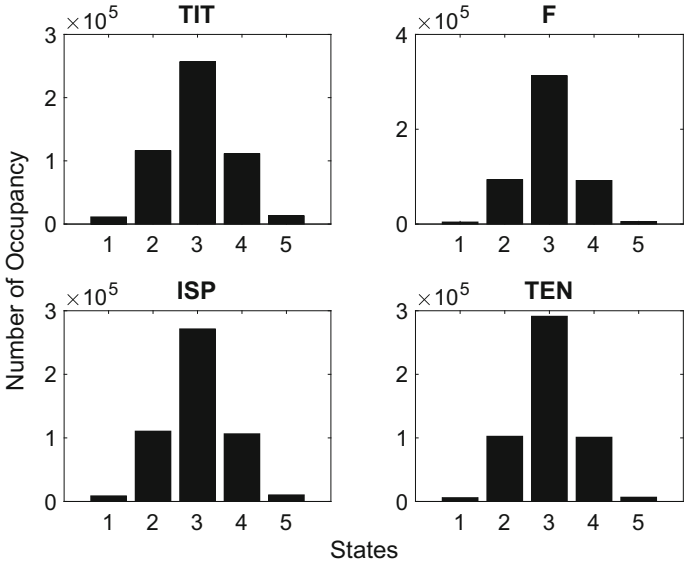

Fig. 15.3 Number of occupancy of each discretized state of the logarithm of volume change

Following D'Amico and Petroni [4] we use as definition of the function f^{λ} in (15.1) an exponentially weighted moving average (EWMA) of the squares of $Z^V(t)$ which has the following expression:

$$f^{\lambda}(J_{n-1-k}, T_n, a) = \frac{\lambda^{T_n-a} J^2_{n-1-k}}{\sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} \lambda^{T_n-a}} = \frac{\lambda^{T_n-a} J^2_{n-1-k}}{\sum_{a=1}^{T_n} \lambda^a}. \tag{15.11}$$

Consequently the index process becomes $I_n^V(\lambda) = \sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_{n-k}-1} \left(\frac{\lambda^{T_n-a} J^2_{n-1-k}}{\sum_{a=1}^{T_n} \lambda^a} \right)$.

The index $I_n^V(\lambda)$ was also discretized into 5 states of low, medium low, medium, medium high and high volume variation. Using these definitions and discretizations we estimated, for each stock, the probabilities defined in the previous section by using their estimators directly from real data. By means of Monte Carlo simulations we were able to produce, for each of the 4 stocks, a synthetic time series.

Each time series is a realization of the stochastic process described in the previous section with the same time length as real data. Statistical features of these synthetic time series are then compared with the statistical features of real data. In particular, we tested our model for the ability to reproduce the autocorrelation functions of the absolute value of $Z^V(t)$ and the first passage time distribution.

We estimated $\Sigma(\tau)$ (see Eq. (15.9)) for real data and for synthetic data and show in Fig. 15.4 a comparison between them for all stocks. The figures show two main results: first of all, as already noticed in Figs. 15.1 and 15.2, the volume process is

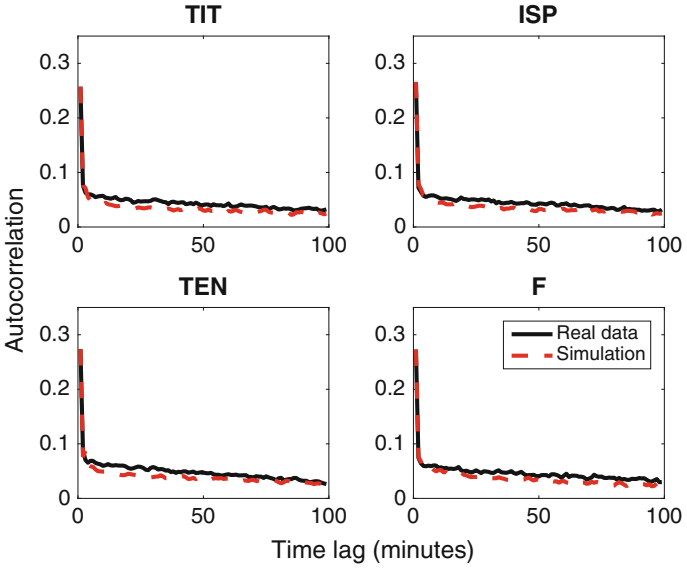

Fig. 15.4 Autocorrelation functions of the absolute value of $Z^V(t)$ for real data (solid line) and synthetic (dashed line) time series

Table 15.2 Percentage square root mean error between real and synthetic autocorrelation function reported in Fig. 15.4

Stock	Error (%)
F	3.6
ISP	3.2
TIT	3.0
TEN	3.9

autocorrelated for a long time, the second results is that the WISMC process is able to capture completely the autocorrelation structures. For each stock we estimated the percentage root mean square error (PRMSE) the results are reported in Table 15.2.

The PRMSE demonstrates what seen in Fig. 15.4, i.e. that our model is able to reproduce almost perfectly the autocorrelation of the absolute value of the logarithmic volume change for each stock. For each of stock in our database we estimate also the first passage time distribution (fpt) as defined in Eq. (15.10) directly from the data (real data) and from the synthetic time series generated as described above.

From Fig. 15.5 it is obvious that the fpt estimated from a WISMC model show exactly the same behaviour of the pft estimated from real data.

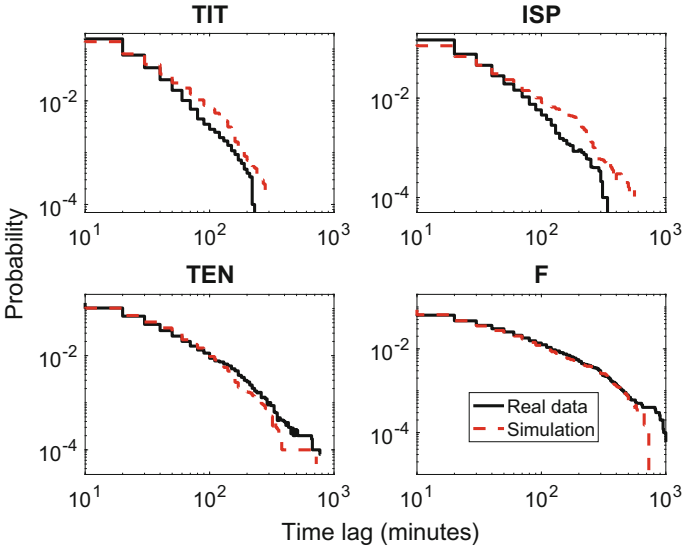

Fig. 15.5 First passage time distribution for $\sigma = 1000$

15.5 Conclusions

In this paper we advanced the use of Weigthed-Indexed Semi-Markov Chain models for modeling high frequency financial volumes. We applied the model on real financial data and we shown that the model is able to reproduce important statistical fact of financial volumes as the autocorrelation function of the absolute values of volumes and the first passage time distribution. Further developments will be a more extensive application to other financial stocks and indexes and the proposal of a complete model where returns, volumes and durations are jointly described.

References

1. Brownlees, C.T., Cipollini, F., Gallo, G.M.: Intra-daily volume modeling and prediction for algorithmic trading. J. Financ. Econ. **9**(3), 489–518 (2011)
2. D'Amico, G., Petroni, F.: A semi-Markov model with memory for price changes. J. Stat. Mech. Theory Exp. P12009 (2011)
3. D'Amico, G., Petroni, F.: Weighted-indexed semi-Markov models for modeling financial returns. J. Stat. Mech. Theory Exp. P07015 (2012)
4. D'Amico, G., Petroni, F.: A semi-Markov model for price returns. Physica A **391**, 4867–4876 (2012)
5. D'Amico, G., Petroni, F.: Multivariate high-frequency financial data via semi-Markov processes. Markov Process. Relat. Fields **20**, 415–434 (2014)
6. D'Amico, G., Petroni, F.: Copula based multivariate semi-Markov models with applications in high-frequency finance. Eur. J. Oper. Res. **267**(2), 765–777 (2018)

7. D'Amico, G., Guillen, M., Manca, R.: Semi-Markov disability insurance models. Commun. Stat. Theory Methods **42**(16), 2872–2888 (2013)
8. De Jong, F., Rindi, B.: The Microstructure of Financial Markets. Cambridge University Press, Cambridge (2009)
9. Fodra, P., Pham, H.: Semi-markov model for market microstructure. Appl. Math. Finan. **22**(3), 261–295 (2015)
10. Gopikrishnan, P., Plerou, V., Gabaix, X., Stanley, H.E.: Statistical properties of share volume traded in financial markets. Phys. Rev. E **62**, 4493–4496 (2000)
11. Jain, P., Joh, G.: The dependence between hourly prices and trading volume. J. Financ. Quant. Anal. **23**, 269–283 (1988)
12. Lobato, I.N., Velasco, C.: Long memory in stock-market trading volume. J. Bus. Econ. Stat. **18**, 410–427 (2000)
13. Mainardi, F., Raberto, M., Gorenflo, R., Scalas, E.: Fractional calculus and continuous-time finance II: the waiting-time distribution. Physica A **287**, 468–481 (2000)
14. Manganelli, S.: Duration, volume and volatility impact of trades. J. Financ. Mark. **8**, 377–399 (2005)
15. Plerou, V., Stanley, H.E.: Tests of scaling and universality of the distributions of trade size and share volume: evidence from three distinct markets. Phys. Rev. E **76**, 46–109 (2007)
16. Podobnik, B., Horvatic, D., Petersen, A.M, Stanley, H.E.: Cross-correlations between volume change and price change. PNAS **29**(52), 22079–22084 (2009)
17. Raberto, M., Scalas, E., Mainardi, F.: Waiting-times and returns in high-frequency financial data: an empirical study. Physica A **314**, 749–755 (2002)
18. Silvestrov, D., Manca, R., Silvestrova, E.: Computational algorithms for moments of accumulated Markov and semi-Markov rewards. Commun. Stat. Theory Methods **43**(7), 1453–1469 (2014)
19. Stenberg, F., Manca, R., Silvestrov, D.: An algorithmic approach to discrete time non-homogeneous backward semi-Markov reward processes with an application to disability insurance. Methodol. Comput. Appl. Probab. **9**(4), 497–519 (2007)
20. Vassiliou, P.-C.G., Papadopoulou, A.: Non-homogeneous semi-Markov systems and maintainability of the state sizes. J. Appl. Probab. **29**, 519–534 (1992)

Chapter 16
PageRank in Evolving Tree Graphs

Benard Abola, Pitos Seleka Biganda, Christopher Engström, John Magero Mango, Godwin Kakuba and Sergei Silvestrov

Abstract In this article, we study how PageRank can be updated in an evolving tree graph. We are interested in finding how ranks of the graph can be updated simultaneously and effectively using previous ranks without resorting to iterative methods such as the Jacobi or Power method. We demonstrate and discuss how PageRank can be updated when a leaf is added to a tree, at least one leaf is added to a vertex with at least one outgoing edge, an edge added to vertices at the same level and forward edge is added in a tree graph. The results of this paper provide new insights and applications of standard partitioning of vertices of the graph into levels using breadth-first search algorithm. Then, one determines PageRanks as the expected numbers of random walk starting from any vertex in the graph. We noted that time complexity of the proposed method is linear, which is quite good. Also, it is important to point out that the types of vertex play essential role in updating of PageRank.

Keywords Breadth-first search · Forward edge · PageRank · Random walk · Tree

B. Abola (✉) · P. S. Biganda · C. Engström · S. Silvestrov
Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Box 883, 72123, Västerås, Sweden
e-mail: benard.abola@mdh.se

C. Engström
e-mail: christopher.engstrom@mdh.se

S. Silvestrov
e-mail: sergei.silvestrov@mdh.se

B. Abola · J. M. Mango · G. Kakuba
Department of Mathematics, School of Physical Sciences, Makerere University, Box 7062, Kampala, Uganda
e-mail: mango@cns.mak.ac.ug

G. Kakuba
e-mail: godwin.a.kakuba@gmail.com

P. S. Biganda
Department of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam, Box 35062, Dar es Salaam, Tanzania
e-mail: pitos.biganda@mdh.se

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_16

16.1 Introduction

PageRank is a measure of Web page quality according to their relative importance while taking into account a hyperlink graph [4]. Since the original work of Brin and Page, ranking web pages in network structure has received considerable attention in the scientific community as pointed out in [1, 8, 9, 14]. The growing number of web pages or vertices in complex networks is one of well known challenge. It is desirable to improve PageRank algorithm to put up with the increasing size of networks while maintaining the requirement for ranks quality as suggested by Brin and Page [4].

Recently, PageRank is used in applications far beyond its origins in Google's web-search. In fact, applications of PageRank has turned out to be much more general and can be applied to numerous types of graph or network. For instance, it has been adopted in bibliometrics, social, road, biological and information network analysis [8].

Methods and ideas similar to PageRank are available, for instance, EigenTrust algorithm [13], applied to reputation management in peer-to-peer networks, citation ranking [18], DeptRank algorithm, which is used to evaluate risk in financial networks [3], and GeneRank, used in microarray data analysis where one measures whether or not a gene's expression is promoted or repressed [17]. Also, in chemistry PageRank algorithm can be applied to study molecules as for example in [16], where in particular, the authors used the algorithm to investigate changes in network of molecules linked to hydrogen bonds among water molecules. From these few usage of PageRank, it is vividly clear that applications and ideas of PageRank has turned out to be stimulating and created new direction in Mathematics and development of algorithms in static or evolving networks as suggested by Engström and Silvestrov [6].

Numerous attempts have been made to speed up PageRank algorithm. For instance, Ishii et al. [12] have aggregated webpages that are close and are expected to have similar PageRank. Engström and Silvestrov [6] have proposed an algorithm that partition the graph into components, that is, strongly connected components (SCC) or connected acyclic component (CAC) and re-calculate PageRank as the network changes. Another method is to remove dangling pages (pages with no links to other pages), and then calculate their ranks at the end [2, 15].

The method proposed here have some similarities with the method proposed by Engström and Silvestrov [7]. The main difference is that we emphasized updating or re-calculating PageRanks of evolving tree graphs without resorting to iterative methods such as Power Method and Jacobi method. Therefore, this article can be considered as a contribution towards PageRank computation in a specific network system (tree graphs), avoiding numerical iterative methods for linear systems.

This article is organized as follows. In Sect. 16.2 we set up definitions, notations and essential concepts. In Sect. 16.3, we present the different types of changes and how to re-calculate PageRanks associated to the changes. Section 16.4 is devoted to analysis of time complexity of the proposed method and finally Sect. 16.5 present conclusions of the study.

16.2 Preliminaries

This section describes important notations and definitions that are used throughout the article.

- T_G: The tree graph consisting of nodes and links for which we want to calculate PageRank. It contains both the system matrix A_G and a weight vector $\vec{v}_G$. A subindex G can be either a capital letter or a number in the case of multiple systems.
- c: A parameter $0 < c < 1$ for calculating PageRank, usually $c = 0.85$.
- T: Global tree graph made up of multiple disjoint subsystems $T = T_1 \cup T_2 \cdots \cup T_N$, where N is the number of forests.
- $P_{\to j} = w_j + \sum_{v_i \in V, v_i \neq v_j} w_i P_{ij}$, where w_j is the weight of vertex v_j and P_{ij} is the hitting probability from v_i to v_j in a random walk on the graph containing the vertices V. The type of random walk we will work with is described in Definition 16.1.
- $P_{ab}(c)$ is the probability to reach v_b starting in v_a after passing through v_c at least once.
- $P_{ab}(\bar{c})$ is the probability to reach v_b starting in v_a without ever passing via v_c.
- PageRank can also be written as: $R_a = \frac{P_{\to a}}{1 - P_{aa}}$

PageRank can be defined in various versions, but we consider non-normalised traditional algorithm which views rank using a random walk on link structure of a graph.

Definition 16.1 Consider a random walk on a graph described by A_G, which is the adjacency matrix weighted such that the sum over every non-zero row is equal to one. In each step with probability $c \in (0, 1)$, move to a new vertex from the current vertex by traversing a random outgoing edge with probability equal to the corresponding edge weight. With probability $1 - c$ or if the current vertex have no outgoing edges, we stop the random walk. The PageRank $\vec{R}$ for a single vertex v_j can be written as

$$R_j = \left(\sum_{v_i \in V, v_i \neq v_j} w_i P_{ij} + w_j \right) \left(\sum_{k=0}^{\infty} (P_{jj})^k \right), \tag{16.1}$$

where P_{ij} is the probability to hit node v_j in the random walk starting in node v_i, before stopping of this random walk. This can be seen as the expected number of visits to v_j if we do multiple random walks, starting in every node once and weighting each of these random walks by the vector of vertex weights $\vec{w}$, [7]. Note that we have both vertex and edge weights and that both the initial and final vertex are counted as being visited in such a random walk (but not counted extra).

Another common approach to obtain PageRank is to rewrite it as an eigenvector problem. This way, one resorts to solving a linear system of the form $(\mathsf{I} - c\mathsf{A}_G^\top)\vec{R} = \vec{\bar{v}}$, Hence, the PageRank of vertex v_j can be expressed as $R_j = [(\mathsf{I} - c\mathsf{A}_G^\top)^{-1}\vec{\bar{v}}]_j$.

16.2.1 Tree Graphs and Partitioning

Definition 16.2 A tree $T(V, E)$ is a connected and acyclic graph, where V and E are sets of vertices and edges respectively. A root of T is the only node without a parent, denoted as r. If two or more nodes have the same parent, then they are called siblings. A node with no children is a leaf. A non-leaf node is termed as internal node or simply called a node.

Definition 16.3 A level in a tree T, is collection of all vertices at a fixed distance to the root. Thus, if a root is at level L, then the children of the root will be at a lower level, say $L - 1$, [10].

Next, we present partitioning of tree graph into levels. Consider a graph $T(V, E)$, with a path length l, that is, a sequence of l edges

$$\{(v_l, v_{l-1}), (v_{l-1}, v_{l-2}), \ldots, (v_1, v_0)\}$$

where all the vertices $v_l, v_{l-1}, \ldots, v_1, v_0$ are distinct. Abbreviate the level of vertex v_l as l, v_{l-1} as $l - 1$ and so on.

If one performs breadth-first search (BFS), the algorithm finds the neighbor of a vertex v_L before finding the neighbor of v_{L-1}, this way a breath-first tree is generated. In general, suppose L_k is the level of vertex $v_k \in V$, then a partition of T by level is understood as the collection of subsets of vertices $v_k \in V$, denoted by L_k such that $L_j \cap L_k = \emptyset$ for $j \neq k$ and $V = L_l \cup L_{l-1} \cup, \ldots, \cup L_1 \cup L_0$. Here we adopt a monotone ordering l for a breadth-first tree, where a vertex is numbered before its father as in [6].

16.2.2 PageRank After Addition or Removal of Edges Between Vertices

By decomposing PageRank for a vertex v_i into two parts depending on if a random walk visited the source vertex v_a (the vertex to which we add or remove edges) before the first visit to v_i or not, we can decompose PageRank as in Lemma 16.1. The result is helpful in determining ranks when edges are added or removed between two vertices. In particular, adding or removing edges has one main effect of outgoing edges from a source vertex, that is, update of the edge weights and levels of target vertices.

We look at this changes in term of weights change on all old outgoing edges of source vertex in the next section.

Lemma 16.1 *PageRank of a single vertex v_b after removing all outgoing edges from vertex v_a or adding new outgoing edges from v_a which previously had no outgoing edges can be written as:*

$$R_b = \frac{P_{\to b}(\bar{a})}{1 - P_{bb}(\bar{a})} \pm \frac{R_a P_{ab}(\bar{a})}{1 - P_{bb}(\bar{a})}, \tag{16.2}$$

where positive + and negative − signs are associated to changes due to addition of new edges and removal of all edges out of source vertex, respectively.

The proof can be found in [7].

16.2.3 PageRank After Rank-One Perturbation of Weighted Adjacency Matrix

Some related results concerning perturbed Markov chains can be found in [11]. A recent application in PageRank computation for the Google web search engine was studied by Xiong and Zheng [19]. They observed that Google matrix is special rank-1 updated (perturbed) matrix whose eigenvector (PageRank) corresponds to the maximal eigenvalue 1. Thus, perturbation of a matrix is in tandem with addition or removal of edges from a source vertex as in Sect. 16.2.2.

This subsection is devoted to determining PageRank vector of the new weighted matrix, $\mathsf{P}^{(2)}$ obtained after perturbation. We claim that by using Sherman–Morrison formula one can obtain explicit formula of evolving network. To ensure that let $\mathsf{P}^{(1)}$ denote unperturbed weighted matrix such that

$$\mathsf{P}^{(2)^\top} = \mathsf{P}^{(1)^\top} + \vec{u}\vec{v}^\top, \tag{16.3}$$

where $\vec{u}\vec{v}^\top \in \mathbb{R}^{n\times n}$ are column rank-one matrices. The PageRank vector $\vec{R}^{(2)}$, of $\mathsf{P}^{(2)^\top}$ can be obtained by solving the classical PageRank linear equation of the form $(\mathsf{I} - c\mathsf{A}_G^\top)\vec{x} = \vec{b}$, and we get

$$\vec{R}^{(2)} = [(\mathsf{I} - c\mathsf{P}^{(1)^\top}) - c\vec{u}\vec{v}^\top]^{-1}\vec{b}, \tag{16.4}$$

where $\vec{b}$ is a column vector of ones. Applying Sherman–Morrison formula as cited in Deng [5] to $[(\mathsf{I} - c\mathsf{P}^{(1)^\top}) - c\vec{u}\vec{v}^\top]^{-1}$, an explicit estimate of $\vec{R}^{(2)}$ can be found.

Lemma 16.2 *Let $G \in \mathbb{R}^{n\times n}$ be a linear operator, matrix $\mathsf{B} \in \mathbb{R}^{n\times n}$ and vectors $\vec{u}, \vec{v} \in \mathbb{R}^{n\times 1}$. Suppose $G = \mathsf{B} - \vec{u}\vec{v}^\top$, then the inverse G^{-1} of G can be expressed as*

$$G^{-1} = \mathsf{B}^{-1} + \frac{\mathsf{B}^{-1}\vec{u}\vec{v}^\top\mathsf{B}^{-1}}{1 - \vec{v}^\top\mathsf{B}^{-1}\vec{u}}, \tag{16.5}$$

where $det(\mathsf{B}) \neq 0$ and $det(\mathsf{B} - \vec{u}\vec{v}^\top) \neq 0$.

Proof Let $\vec{x}, \vec{b} \in \mathbb{R}^n$ be such that $G\vec{x} = \vec{b}$. Then from $G = \mathsf{B} - \vec{u}\vec{v}^\top$ we get

$$
\begin{aligned}
(\mathsf{B} - \vec{u}\vec{v}^{\top})\vec{x} &= \vec{b}, \\
\mathsf{B}\vec{x} - \vec{u}\vec{v}^{\top}\vec{x} &= \vec{b}, \\
\mathsf{B}\vec{x} &= \vec{b} + \vec{u}\vec{v}^{\top}\vec{x}, \\
\vec{x} &= \mathsf{B}^{-1}\vec{b} + \mathsf{B}^{-1}\vec{u}\vec{v}^{\top}\vec{x}. \qquad (16.6)
\end{aligned}
$$

Pre-multiplying $\vec{x}$ by $\vec{v}^{\top}$ and collecting the term containing $\vec{v}^{\top}\vec{x}$ yields

$$(\mathsf{I} - \vec{v}^{\top}\mathsf{B}^{-1}\vec{u})\vec{v}^{\top}\vec{x} = \vec{v}^{\top}\mathsf{B}^{-1}\vec{b}$$

From here we obtain $\vec{v}^{\top}\vec{x} = \frac{\vec{v}^{\top}\mathsf{B}^{-1}\vec{b}}{\mathsf{I}-\vec{v}^{\top}\mathsf{B}^{-1}\vec{u}}$ and substituting in (16.6) we get

$$\vec{x} = \left(\mathsf{B}^{-1} + \frac{\mathsf{B}^{-1}\vec{u}\vec{v}^{\top}\mathsf{B}^{-1}}{1 - \vec{v}^{\top}\mathsf{B}^{-1}\vec{u}}\right)\vec{b}.$$

Hence, $G^{-1} = \mathsf{B}^{-1} + \frac{\mathsf{B}^{-1}\vec{u}\vec{v}^{\top}\mathsf{B}^{-1}}{1-\vec{v}^{\top}\mathsf{B}^{-1}\vec{u}}$ as required. □

Remark 16.1 We wish to emphasize that matrix perturbation approach only addresses change due to addition or removal of edges but not for vertices.

16.3 PageRank of Evolving Tree Graphs

The effect of changes involving only few vertices in large sparse chains such as in Google's PageRank application is primarily local and most PageRanks are not affected significantly [7, 14]. Thus, it is important to focus on some parts of the graph that have changed. Obviously, some changes may not require update of the previous ranks while other do. For instance, adding a new leaf to a vertex without outgoing edge(s) does not require updating the previous ranks. One needs to calculate the PageRank of the new vertices only. Since there are many changes that occur, we restrict ourselves to four categories which are presented in the sub-sections that follows.

16.3.1 Adding a Leaf to a Tree

Consider a directed tree T with four vertices n_1, n_2, n_3 and n_4, suppose a new vertex e_1 is linked to n_4 as shown in Fig. 16.1. If we assume that n_4 has no outgoing edge before adding node e_1, then the PageRank of vertices remain the same after the addition. Whereas, the PageRank R_{e_1} of e_1, becomes $w_{e_1} + c_{n_4 \to e_1} R_{n_4}$, where w_{e_1} is the weight of e_1 and $c_{n_4 \to e_1}$ is 1-step probability to go from n_4 to e_1. Alternatively, one needs to generate breadth first tree of the new tree graph $T \cup e_1$ using BFS algorithm. Hence, vertex e_1 will be at the lowest level (level 0) and its PageRank can

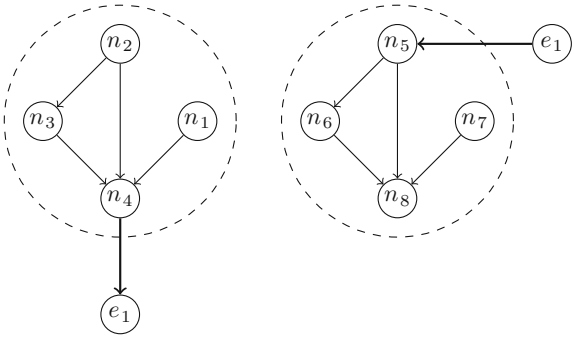

Fig. 16.1 Adding a vertex e_1 to a tree originally having $V = \{n_1, n_2, n_3, n_4\}$

easily be calculated when the PageRank of level 1 and above are known. Precisely, the PageRank of parents or ancestors influence those of the children. In summary, any change on a vertex at the lower can be considered as localised perturbation which requires updating PageRank of smaller subgraph. We state a corollary that will generalized this kind of re-calculating PageRank after adding a leaf to a tree.

Proposition 16.1 *Given PageRank R_a of vertex v_a, where v_a has no outgoing edge, and adding a leaf at the vertex v_a such that there is 1-step transition probability from v_a to v_b denoted by c_{ab}, the PageRank of the leaf v_b can be expressed as*

$$R_b = w_b + c_{ab} R_a, \tag{16.7}$$

where w_b is the weight of v_b.

Proof Using Definition 16.1 and rewriting the PageRank of vertex v_b as $P_{\to b} = w_b + \sum_{v_i \in V, v_i \neq v_b} w_i P_{ib}$, the term $\sum_{v_i \in V, v_i \neq v_b} w_i P_{ib} = w_a P_{ab}$ since there is no other path from v_a to v_b. It can be seen that $w_a = R_a$ and $P_{ab} = c_{ab}$, hence $P_{\to b} = w_b + c_{ab} R_a$. □

Proof Alternatively, consider PageRank of vertex v_b as rank-one perturbation of weighted adjacency matrix, where an edge is added between v_a and v_b. Assume that the PageRank $\vec{R}^{(1)}$ of unperturbed matrix $\mathsf{P}^{(1)}$ is known. We use of Lemma 16.2. First, fix $\vec{x} = \vec{R}^{(2)}$, $\mathsf{B}^{-1} = (\mathsf{I} - c\mathsf{P}^{(1)^\top})^{-1}$, $\vec{p} = \vec{u}$, $\vec{q} = \vec{v}$ and $\vec{b} = \vec{w}$. Secondly, suppose that $\vec{u}$ and $\vec{v}^\top$ are the b^{th} column vector of identity matrix with the size as $\mathsf{P}^{(1)}$ or $\mathsf{P}^{(2)}$ and row vector of $\mathsf{P}^{(2)}$ respectively. Recall that $\vec{R}^{(1)} = [(\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}\vec{w}$. Then $\vec{R}^{(2)}$ becomes

$$\begin{aligned}\vec{R}^{(2)} &= [(\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}\vec{w} + c\frac{[\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}\vec{u}\vec{v}^\top[\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}}{1 - \vec{v}^\top[\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}\vec{u}}\vec{w}, \\ &= \vec{R}^{(1)} + \frac{[\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}[c\vec{u}\vec{v}^\top]\vec{R}^{(1)}}{1 - \vec{v}^\top[\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}\vec{u}},\end{aligned} \tag{16.8}$$

where $[\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}[c\vec{u}\vec{v}^\top]$ can be seen as a transition matrix from source to target vertex that has zeros everywhere except at entry (v_b, v_a) and also

$$\vec{v}^\top \left[\mathsf{I} - c\mathsf{P}^{(1)^\top}\right]^{-1} \vec{u} = P_{bb}(\bar{a})$$

is the probability from v_b to v_b without going through v_a. Now, let $\vec{R}_b^{(2)}$ and $\vec{R}_b^{(1)}$ be PageRank of the target vertex v_b before and after the change respectively, then (16.8) becomes

$$\vec{R}_b^{(2)} = \vec{R}_b^{(1)} + \frac{c\vec{R}_a^{(1)}}{1 - P_{bb}(\bar{a})}.$$

Since there is no link to reach v_b except from v_a then $P_{bb}(\bar{a}) = 0$, $R_b^{(1)} = w_b$ and $R_b^{(2)} = R_b$ hence we get $R_b = w_b + cR_a^{(1)}$. □

Example 16.1 Consider a graph in Fig. 16.1 with matrices $\mathsf{P}^{(1)}$ and $\vec{u}\vec{v}^\top$ corresponding to the weighted adjacency matrices before addition of new edge $v_{n_4} \to v_{e_1}$ and evolving term respectively. For the purpose of consistency, let the source vertex $v_a = v_{n_4}$ and target vertex $v_b = v_{e_1}$.

Define

$$c\mathsf{P}^{1^\top} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{c}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \frac{c}{2} & c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \vec{u} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{v}^\top = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

and assume that the PageRank of $\mathsf{P}^{(1)}$ is

$$\begin{pmatrix} R_{n_2}^{(1)} \\ R_{n_3}^{(1)} \\ R_{n_1}^{(1)} \\ R_a^{(1)} \\ R_b^{(1)} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 + \frac{c}{2} \\ 1 \\ 1 + \frac{3c}{2} + c(1 + \frac{c}{2}) \\ 1 \end{pmatrix}.$$

Using (16.8), we get

$$[\mathsf{I} - c\mathsf{P}^{(1)^\top}]^{-1}[c\vec{u}\vec{v}^\top] = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c & 0 \end{pmatrix} \begin{pmatrix} R_{n_2}^{(1)} \\ R_{n_3}^{(1)} \\ R_{n_1}^{(1)} \\ R_a^{(1)} \\ R_b^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ c\vec{R}_a^{(1)} \end{pmatrix}$$

By Proposition 16.1, we get

$$R_b^{(2)} = 1 + c\underbrace{\left[1 + \frac{3c}{2} + c\left(1 + \frac{c}{2}\right)\right]}_{R_a^{(1)}}.$$

This result is rather interesting. It is apparent that if the previous PageRank is known and a leaf is added to a vertex with no outgoing edge previously, then the PageRank of the graph after the change can be obtained in at most constant time. If we look the updating procedure in term of time complexity then the proposed method save users from computing the new PageRank from scratch.

Following Fig. 16.1 on the right, where a leaf is added to a tree, we explain PageRank update without giving explicit formula for such changes. The PageRank can be addressed as follows:

(i) If the target vertex has no outgoing edge then target vertex is update according to proposition one.
(ii) If the target vertex has at least one outgoing edge then update the target vertex according to proposition one and it corresponding children sequentially.

Extending this idea to a vertex with at least one outgoing edge in the next subsection. Let us begin by presenting without proof a lemma for changes in personalization vector [6].

Lemma 16.3 *Consider a graph with PageRank* $\vec{R}^{(1)}$ *and weight vector* $\vec{w}^{(1)}$. *The new PageRank* $\vec{R}^{(2)}$ *with a new personalization vector* $\vec{w}^{(2)} = \vec{w}^{(1)} + \triangle\vec{w}^{(1)}$, *after adding a set of new outgoing edges from* v_a *can be written as:*

$$R_j^{(2)} = R_j^{(1)} + \left(\triangle w_j^{(1)} + \sum_{v_i \in S, v_i \neq v_j} \triangle w_i^{(1)} P_{ij}\right) \sum_{j=0}^{\infty} P_{jj}. \tag{16.9}$$

16.3.2 Adding k-Leaf to a Vertex with at Least One Outgoing Edge

Let $T_1 = G(V, E)$ be a tree graph such that $V = \{n_1, n_2, n_3, n_4, e_1\}$ and $T_2 = T_1 \cup e_2$, where e_2 is a new leaf as shown in the Fig. 16.2. Take n_4 as a source vertex. Then we observe the following: (1) the PageRank of n_4 and those vertices at level greater are unchanged (2) The weight of e_1 needs to be updated and the level unchanged. We know that n_1 and n_2 have outgoing edges only, then applying Definition 16.1, $R_{n_1}^{(1)} = R_{n_2}^{(1)} = 1$. Subsequently, $R_3^{(1)} = 1 + \frac{c}{2}R_2^{(1)}$, $R_{n_4}^{(1)} = 1 + cR_{n_1}^{(1)} + \frac{c}{2}R_{n_2}^{(1)} + cR_{n_3}^{(1)}$ and $R_{e_1}^{(1)} = 1 + cR_{n_4}^{(1)}$. Clearly, we note that $R_4^{(1)}$ dependent the PageRank of n_1, n_2 and n_3 which are at higher levels compared to the level of n_4. Similarity, $R_{e_1}^{(1)}$ is influence by PageRank of vertex n_4 and above.

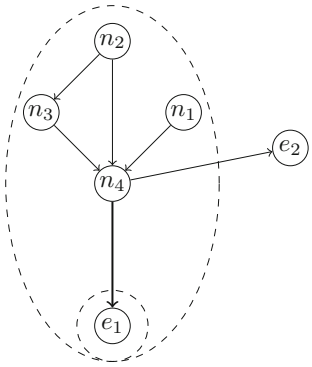

Fig. 16.2 Adding a leaf e_2 to a tree with $V = \{n_1, n_2, n_3, n_4, e_1\}$

Next, let us add a leaf to n_4 to give T_2 and suppose $R^{(2)}$ denotes PageRank after addition of the leaf, then $R^{(2)}_{n_i} = R^{(1)}_{n_i}$, for $i = 1, \ldots, 4$. The rank of e_1, $R^{(2)}_{e_1}$ becomes $1 + \frac{c}{2} R^{(1)}_4$. Rewriting it in term of old PageRank plus update value gives

$$R^{(2)}_{e_1} = R^{(1)}_{e_1} + \underbrace{\left(\frac{1}{2} - 1\right)}_{\triangle w_{n_4}} c R^{(1)}_4 = R^{(1)}_{e_1} + \left(\frac{-c}{2}\right) \vec{R}^{(2)}_4 = 1 + \frac{c}{2} R^{(1)}_4. \tag{16.10}$$

In overall PageRank vector $\vec{R}^{(2)}$ of T_2 is expressed as

$$\begin{pmatrix} R^{(2)}_1 \\ R^{(2)}_2 \\ R^{(2)}_3 \\ R^{(2)}_4 \\ R^{(2)}_{e_1} \\ R^{(2)}_{e_2} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 + \frac{c}{2} \\ 1 + \frac{5c}{2} + \frac{c^2}{2} \\ 1 + \frac{c}{2} + \frac{5c^2}{4} + \frac{c^3}{4} \\ 1 + \frac{c}{2} + \frac{5c^2}{4} + \frac{c^3}{4} \end{pmatrix}.$$

We can see that $R^{(2)}_{e_1} = R^{(2)}_{e_2}$ since they are at the same level and having only a single parent n_4. To explore this case further, we state a theorem that encompass calculating PageRank after addition of outgoing edges to a vertex with outgoing edges previously. Under this consideration, we assume the following:

1. Only outgoing edges are added to a source vertex with at least one such edge before.
2. The addition of edges does not create any cycle in new graph.
3. The level and PageRank of the source vertex and greater are known.

Lemma 16.4 *Consider a tree graph, $T(V, E)$ with PageRank vector $\vec{R}^{(1)}$. Suppose $v_a \in V$ is a source vertex with $m(\geq 1)$ outgoing edges. By adding a set of $k(\geq 1)$ new leaves (outgoing edges) to v_a, the new PageRank $\vec{R}^{(2)}$ of v_a or v_b can be expressed as:*

$$R_a^{(2)} = R_a^{(1)}, \quad R_b^{(2)} = R_b^{(1)} + \left(\frac{m}{n} - 1\right) P_{ab}^{(1)} R_a^{(2)}, \tag{16.11}$$

where $n = m + k$ and $P_{ab}^{(1)}$ is the 1-step transition probability from $v_a \to v_b$.

Proof We start by calculating the PageRank of the nodes v_a on $T(V, E)$, since we had assumed that the PageRank of the source vertex remains the same after adding outgoing edges, then $R_a^{(2)} = R_a^{(1)}$.

To proof the second part, we apply Lemma 16.3. Let $\triangle w_j = 0$, $\triangle w_a^{(1)} = w_a^{(2)} - w_a^{(1)} = (\frac{1}{n} - \frac{1}{k}) R_a^{(2)}$ and $\sum_{l=0}^{\infty} (P_{bb})^l = 1$. Also, there is only a 1-step transition probability from $v_a \to v_b$, then $\sum_{v_i \in S, v_i \neq v_j} \triangle w_i^{(1)} P_{ij} = \triangle w_a^{(1)} P_{ab}^{(1)}$. Substituting the known terms in the formula (16.9)

$$R_j^{(2)} = R_j^{(1)} + \left(\triangle w_j^{(1)} + \sum_{v_i \in S, v_i \neq v_j} \triangle w_i^{(1)} P_{ij} \right) \sum_{l=0}^{\infty} (P_{bb})^l.$$

we get $R_b^{(2)} = R_b^{(1)} + (\frac{m}{n} - 1) P_{ab}^{(1)} R_a^{(2)}$. □

16.3.3 Adding or Removing Edges from Vertices at the Same Level

In all cases considered previously we first perform BFS algorithm to obtain breadth-first tree before determining rank of vertices. Also, the vertices of the tree are labelled according to their discovery and each vertex is assigned a label with corresponding level. We noticed that vertex at higher level influence the PageRank of those at the lower level. Thus, any change of internal edge influence the breadth-first tree structure which depend on edge classification as mentioned by Yan and Han [20]. Here we focus on one classification, forward edge. This refers to those edges that describes ancestor-to-descendant relation, that is, high to low level set of vertices. Another form of edge class examined in this article is an edge linked to vertices at the same level. We explored the influence of the two classes in PageRanks re-calculation.

Further, vertex classification contributes to the way we compute PageRank in a graph [6]. Here, we define four distinct classes or groups and it should be emphasized that every vertex in a tree graph belong to single group.

Definition 16.4 For the vertices of a simple directed graph without cycle, we can define four distinct groups of vertices.

1. G_1: vertices with no outgoing or incoming edges.

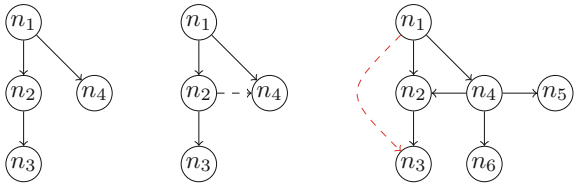

Fig. 16.3 Addition of edges between vertices in tree graphs

2. G_2: vertices with no outgoing edges and at least incoming edges (also dangling vertices /leaves.)
3. G_3: vertices with at least one outgoing edge but no incoming edges (also called root vertices)
4. G_4: vertices with at least one out going and incoming edges.

We briefly describe how changes based on edge classification influence Pagerank calculation.

- *Edge added to vertices at the same level*: In reference to Fig. 16.3 (left), it can be seen that $n_2 \in G_4$ and $n_4 \in G_2$ are at the same level. Suppose an edge (n_2, n_4) is added to give middle figure, then update the level and PageRank of n_4. Also, the level and PageRank of vertices below n_3 and n_4 need to be updated.

The following lemma encompasses the re-calculation of PageRank associated to this form of change. For proof see Proposition 16.1 and Lemma 16.4.

Lemma 16.5 *Let $R_a^{(1)}$ and $R_b^{(1)}$ be the PageRanks of vertices v_a and v_b respectively. Suppose that the vertices are previously at the same level L. Then the PageRank $R_b^{(2)}$ after adding an edge (v_a, v_b) can be expressed as $R_b^{(2)} = R_b^{(1)} + c_{ab} R_a^{(2)}$, where c_{ab} is 1-step probability of a random work from vertex v_a to v_b. Also, the Pagerank of vertices linked to v_a only can be updated as $R_{a^-}^{(2)} = R_{a^-}^{(1)} + (\frac{1}{n} - 1) R_a^{(1)}$, where $n = m + 1$, and a^- denotes the vertices at level $L - 1$ linked to v_a before change.*

- *Forward edge* (indicted in red in Fig. 16.3 (right)): Update all the levels and PageRanks of $T - n_1$ sequentially.

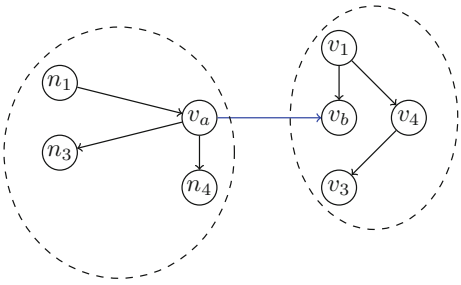

Fig. 16.4 Adding a tree to a tree when the vertices v_a and v_b in G_4 and G_2 groups respectively

16.3.4 Adding Tree to Tree

We now present the influence of adding an edge between trees. The focus has been on whether levels and Pageranks should be updated or not when two-tree graphs are added. We based our description on vertex grouping as pointed in earlier. Consider Fig. 16.4, where $v_a \in T_1$ and $v_b \in T_2$ belong to G_4 and G_2 respectively. It can be seen that PageRank of n_1, v_a, v_1, v_3, v_4 remain but their need to be updated. On the other hand PageRanks and levels of n_3, n_4, v_b require update. Some change in networks may not require overall calculation of PageRank of a system from the scratch but updating few vertices only. Such benefit should be exploited if we are to speed up computation of Pagerank in large sparse graph. Table 16.1 describes a strategy to update to handle PageRank computation based on vertex classification.

Table 16.1 Updating levels and PageRank after adding an edge between two and tree. NB: v_a and v_b are source and target vertices respectively

From T_1 vertex	To T_2 vertex	Description
G_4	G_1	• The PageRank of vertices in T_1 not affected except the outgoing vertex at v_a to be updated
		• Update the Pagerank of vertex v_b only in T_2
		• Update the level of the new tree graph
	G_2	• The PageRank of vertices in T_1 not affected except the linked vertex v_a to be updated
		• Update the vertex $v_b \in T_2$ and all siblings or descendants
		• Update the level of the new tree graph
	G_3	• Update the PageRank of vertices in T_1
		• Update the vertices with incoming edges from $v_b \in T_2$
		• Update the level of the new tree graph
	G_4	• Update the PageRank of vertices in T_1 which are linked by outgoing edges of v_a including the descendants
		• Update the vertex $v_b \in T_2$ and all siblings or descendants
		• Update the level of the new tree graph
G_3	G_1	• Update all the PageRank of vertices in T_1 except v_a or the root
		• Update the Pagerank of vertex v_b only in T_2
		• Update the level of the new tree graph or forest
	G_2	• Update all the PageRank of vertices in T_1 except the root
		• Update the Pagerank of vertex v_b only in T_2
		• Update the level of the new tree graph
	G_3	• Update the PageRank of vertices in T_1 except v_a
		• Update all the vertices T_2
		• Update the level of the new tree graph

16.4 Analysis of Time Complexity of the Changes

This section presents time complexity that arise due to the four changes encountered in Sect. 16.3. We evaluate on computational complexity of each change to determine if there is any advantage of updating PageRank. For a directed graph if we are to determine whether a vertex u is connected to v amounts to traversing the vertices and edges using breadth first search or depth first search algorithm. Also, it is known that the complexity of both algorithms are the same, that is., $\mathcal{O}(|V| + |E|)$, where $|V|$ and $|E|$ is the number of vertices and edges respectively. Recall that one of the purpose of using BFS algorithm is to partition the vertices into levels. This reordering strategy fits evolving graph because one could explore scenario where addition or removal of edges(s)/vertices is localised change or the whole graph perturbation.

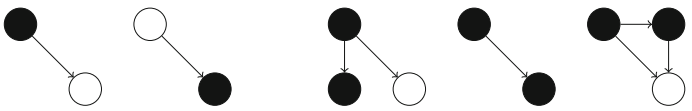

Fig. 16.5 The following figures show some of the main changes considered, labelled as I, II, III, IV, V. The black and hollow nodes represent a tree and a leaf respectively

Table 16.2 Analysis of time complexity of types of change

Case	Description of change	Complexity of the algorithm
I, II	Adding a leaf to a tree	Since we are dealing with directed graph, accessing the target vertex requires one operation only. Hence, time complexity is $\mathcal{O}(1)$. Also, if the target vertex in a tree has no outgoing edge then the complexity equal to that in case I. However, if the vertex has at least one outgoing edge then the time complexity is at most $\mathcal{O}(\|V_1\| + \|E_1\|) \approx \mathcal{O}(E_1)$ if $\|E_1\| \geq \|V_1\|$
III	A leaf is added to a tree with outgoing edges	Obviously, the level of the leaf and a tree below are lower. Hence, we need to update only Pagerank at the lower level. The complexity is equal to $\mathcal{O}(\|V_1 + 1\| + \|E_1 + 1\|) \leq \mathcal{O}(\|V_1\| + \|E_1\|)$. In fact the computational complexity is this change almost the same as in the previous cases
IV	Adding a tree to a tree	Using the same argument as in case II and III. Under a worse case, the complexity is at most $\mathcal{O}(\|V_1\| + \|E_1\|)$. In practice it will be much faster to compute PageRank using the proposed method compared to iterative methods which dependent on convergence or error tolerance
V	Adding an edge from internal vertex to a leaf	In this case, we need to loop through the vertices of a subgraph to ensure that all vertices are appropriately re-labelled. This is done in linear time and discovery of new edge takes $\mathcal{O}(1)$. Hence, we get linear time complexity

As we show earlier, some changes do not require overall re-calculation of PageRank of the graph from the scratch but only updating small component of a graph. Hence, the complexity is less than $\mathcal{O}(|V| + |E|)$. We emphasize that, in comparison to power method of computing PageRank, the cost of overall iteration is about $\mathcal{O}|V|^3$ plus indexing overhead due to storage scheme. Although, BFS algorithm is space bound in many practice, the space complexity is also linear. Thus, it is reasonable to suggest that partition scheme has high advantage in calculating PageRank of graphs. Moreover, the update tackle in this article allows for addition or removal of vertices or edges at will.

To this end, we present analysis of time complexity associated to the specific changes as summarised in Fig. 16.5. The description is highlighted in Table 16.2. We can not exhaust all kind of changes in tree graph because they are practically many.

16.5 Conclusions

In this article, we have shown how it is feasible to update PageRank in evolving tree graph without resorting to iterative methods such as LU decomposition, Jacobi and the like. We began by showing how PageRank can be updated when a leaf is added to a tree. We extended the concept to multiple addition of leaves on a source vertex with at least one outgoing edge. Further, we investigated change in DAG when an edge is added between vertices at the same level.

The main importance of this technique is that it allows us to do changes in both edges and vertices at will. Moreover, some changes require only localised update of PageRank of a network which can be done in linear time as compared to the convention methods of PageRank calculation. In fact, most of iterative methods have computational complexity of $\mathcal{O}|V|^3$ plus overhead due to storage scheme which is quite higher than the proposed method.

Acknowledgements This research was supported by the Swedish International Development Cooperation Agency (Sida), International Science Programme (ISP) in Mathematical Sciences (IPMS), Sida Bilateral Research Program (Makerere University and University of Dar-es-Salaam). We are also grateful to the research environment Mathematics and Applied Mathematics (MAM), Division of Applied Mathematics, Mälardålen University for providing an excellent and inspiring environment for research education and research. The authors are grateful to Dmitrii Silvestrov for useful comments and discussions.

References

1. Anderson, F., Silvestrov, S.: The mathematics of internet search engines. Acta Appl. Math. **104**, 211–242 (2008)
2. Arasu, A., Novak, J., Tomkins, A., Tomlin, J.: PageRank computation and the structure of the web: experiments and algorithms. In: Proceedings of the Eleventh International World Wide Web Conference, Poster Track, pp. 107–117 (2002)

3. Battiston, S., Puliga, M., Kaushik, R., Tasca, P., Calderelli, G.: DeptRank: too central to fail? Financial networks, the fed and systemic risk. Technical report (2012)
4. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Comput Netw ISDN Syst. **30**(1–7), 107–117 (1998) (Proceedings of the Seventh International World Wide Web Conference)
5. Deng, C.Y.: A generalization of the Sherman-Morrison-Woodbury formula. Appl. Math. Lett. **24**(9), 1561–1564 (2011)
6. Engström, C., Silvestrov, S.: A componentwise PageRank algorithm. In: Bozeman, J.R., Oliveira, T., Skiadas, C.H. (eds.) Stochastic and Data Analysis Methods and Applications in Statistics and Demography, ASMDA 2015 Proceedings: 16th Applied Stochastic Models and Data Analysis International Conference with 4th Demographics 2015 Workshop, ISAST: International Society for the Advancement of Science and Technology, pp. 375–388 (2016)
7. Engström, C., Silvestrov, S.: Calculating Pagerank in a changing network with added or removed edges. In: Proceedings of ICNPAA 2016 World Congress, AIP Conference Proceedings 1798, 020052-1-020052-8 (2017). https://doi.org/10.1063/1.4972644
8. Gleich, D.F.: PageRank beyond the Web. SIAM Rev. **57**(3), 321–363 (2015)
9. Gleich, D.F., Gray, A.P., Greif, C., Lau, T.: An inner-outer iteration for computing PageRank. SIAM J. Sci. Comput. **32**(1), 349–371 (2010)
10. Harris, J.M., Hirst, J.L., Mossinghoff, M.J.: Graph Theory. Combinatorics and Graph Theory, pp. 1–127. Springer, Berlin (2008)
11. Hunter, J.J.: Stationary distributions of perturbed Markov chains. Linear Algebra Appl. **82**, 201–214 (1986)
12. Ishii, H., Tempo, R., Bai, E., Dabbene, F.: Distributed randomized PageRank computation based on web aggregation. In: Proceedings of the 48h IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference, pp. 3026–3031 (2009)
13. Kamvar, S.D., Schlosser, M.T., Garcia-Molina, H.: The eigentrust algorithm for reputation management in P2P networks. In: Proceedings of the 12th International Conference on World Wide Web, WWW'03, pp. 640–651 (2003)
14. Langville, A.N., Mayer, C.D.: Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton (2006)
15. Lee, C.P.C., Golub, G.H., Zenios, S.A.: A fast two-stage algorithm for computing PageRank and its extensions. Sci. Comput. Comput. Math. **1**(1), 1–9 (2003)
16. Mooney, B.L., Corrales, L.R., Clark, A.E.: Molecular networks: an integrated graph theoretic and data mining tool to explore solvent organization in molecular simulation. J. Comput. Chem. **33**(8), 853–860 (2012)
17. Morrison, J.L., Breitling, R., Higham, D.J., Gilbert, D.R.: GeneRank: using search engine technology for the analysis of microarray experiments. BMC Bioinform. **6**(1), 233 (2005)
18. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab (1999)
19. Xiong, Z., Zheng, B.: On the eigenvalues of a specially updated complex matrix. Comput. Math. Appl. **57**(10), 1645–1650 (2009)
20. Yan, X., Han, J.: gSpan: graph-based substructure pattern mining. In: Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference, pp. 721–724 (2002)

Chapter 17
Traditional and Lazy PageRanks for a Line of Nodes Connected with Complete Graphs

Pitos Seleka Biganda, Benard Abola, Christopher Engström, John Magero Mango, Godwin Kakuba and Sergei Silvestrov

Abstract PageRank was initially defined by S. Brin and L. Page for the purpose of measuring the importance of web pages (nodes) based on the structure of links between them. Due to existence of diverse methods of random walk on the graph, variants of PageRank now exists. They include traditional (or normal) PageRank due to normal random walk and Lazy PageRank due to lazy random walk on a graph. In this article, we establish how the two variants of PageRank changes when complete graphs are connected to a line of nodes whose links between the nodes are in one direction. Explicit formulae for the two variants of PageRank are presented. We have noted that the ranks on a line graph are the same except their numerical values which

P. S. Biganda (✉) · B. Abola · C. Engström · S. Silvestrov
Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Box 883, 72123 Västerås, Sweden
e-mail: pitos.biganda@mdh.se

C. Engström
e-mail: christopher.engstrom@mdh.se

S. Silvestrov
e-mail: sergei.silvestrov@mdh.se

P. S. Biganda
Department of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam, Box 35062, Dar es Salaam, Tanzania

B. Abola · J. M. Mango · G. Kakuba
Department of Mathematics, School of Physical Sciences, Makerere University, Box 7062, Kampala, Uganda
e-mail: benard.abola@mdh.se

J. M. Mango
e-mail: mango@cns.mak.ac.ug

G. Kakuba
e-mail: godwin.a.kakuba@gmail.com

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_17

differ. Further, we have observed that both normal random walk and lazy random walk on complete graphs spend almost the same time at each node.

Keywords Graph · Random walk · PageRank · Lazy PageRank

17.1 Introduction

PageRank was first introduced by Brin and Page [6] to rank homepages (nodes) on the Internet, based on the structure of links. When a person is interested in getting a certain information from the internet, one is most likely going to use a search engine (eg. Google search engine) to look for such information. Moreover, the person will be interested in getting the most relevant ones. What PageRank aims to do, is to sort out and place the most relevant pages first in the list of all information displayed after the search.

It is known that the number of pages on the internet is very large and keeps on increasing over time. For this reason, the PageRank algorithm need to be very fast to accommodate the increasing number of pages and at the same time retaining the requirement for quality of the ranking results as one carries out an internet search [6].

Algorithms similar to PageRank are available, for instance, EigenTrust algorithm, by Kamvar et al. [16], applied to reputation management in peer-to-peer networks, and DeptRank algorithm, which is used to evaluate risk in financial networks [2]. These imply that PageRank concept can be adopted to various networks problems.

Usually PageRank is calculated using the power method. The method has been found to be efficient for both small and large systems. The convergence speed of the method on a webpage structure depends on the parameter c, where c is a real number such that $0 < c < 1$ [12], and the problem is well conditioned unless c is very close to 1 [14]. However, many methods have been developed for speeding up the calculations of PageRank in order to meet the increasing number of pages on the internet. Some of these methods include aggregating webpages that are close and are expected to have similar PageRank [13], partitioning the graph into components as in [11], removing the dangling nodes before computing PageRank and then calculate their ranks at the end or use a power series formulation of PageRank [1], and not computing the PageRank of pages that have already converged in every iteration as suggested by Sepander et al. [15].

There are also studies on a large scale using PageRank and other measure in order to learn more about the Web. One of them is looking at the theoretical and experimental perspective of the distribution of PageRank as by Dhyani et al. [8].

The theory behind PageRank is built from Perron-Frobenius theory [4] and the study of Markov chains [18]. But how PageRank changes with changes in the system or parameters is not well known. Engström and Silvestrov[9, 10] investigated the changes of PageRank of the nodes in the system consisting of a line of nodes and an outside node, a complete graph connected to the line of nodes and connecting a

simple line with a complete graph where one node in the line is part of the complete graph. In [5], we extended their work by looking at a simple line connected to multiple outside nodes and a simple line connected to two complete graphs. In all the cases studied in this article, we considered traditional (or normal) PageRank as the solution to a linear system of equations and also as probabilities of a random walk through a graph. We develop explicit formulas for the lazy PageRank and compare with those of traditional PageRank developed in our previous work to determine the ranking behaviour as the system changes.

A structure of this article is as follows. In Sect. 17.2, we give some preliminaries which include notations, definitions and some initial results that are essential in this work. The main results are presented in Sects. 17.3 and 17.4. Finally, Sect. 17.5 contains concluding remarks of the work.

17.2 Preliminaries

This section describes important notations and definitions. We start by giving some notations and thereafter essential definitions that are used throughout the article.

- S_G: The system of nodes and links for which we want to calculate PageRank. It contains both the system matrix A_G and a weight vector $\vec{v}_G$. A subindex G can be either a capital letter or a number in the case of multiple systems.
- n_G: The number of nodes in system S_G.
- A_G: A system matrix of size $n_G \times n_G$ where an element $a_{ij} = 0$ means there is no link from node i to node j. Non-zero elements are equal to $1/r_i$ where r_i is the number of links from node i.
- $\vec{u}_G$: Non-negative weight vector, not necessary with sum one. Its size is $n_G \times 1$.
- c: A parameter $0 < c < 1$ for calculating PageRank, usually $c = 0.85$.
- $\vec{g}_G$: A vector with elements equal to one for dangling nodes and zero otherwise in S_G. Its size is $n_G \times 1$.
- M_G: Modified system matrix, $\mathsf{M}_G = c(\mathsf{A}_G + \vec{g}_G\vec{u}_G^{\top})^{\top} + (1-c)\vec{u}_G\vec{e}^{\top}$ used to calculate PageRank $\vec{R}_G$, where $\vec{e}$ is the vector whose all entries are ones. Size $n_G \times n_G$.
- S: Global system made up of multiple disjoint subsystems $S = S_1 \cup S_2 \ldots \cup S_N$, where N is the number of subsystems.

In the cases where there is only one possible system the subindex G is omitted. For systems making up S we define disjoint systems in the following way.

Definition 17.1 Two systems S_1, S_2 are disjoint if there are no paths from any nodes in S_1 to S_2 or from any nodes in S_2 to S_1.

PageRank can be defined in various versions, for instance in [9] two versions were presented. However, in this paper we adopt the non-normalized PageRank, denoted as R_j for node j which are denoted as $R_j^{(t)}$ and $R_j^{(l)}$ for the non-normalized traditional PageRank and lazy PageRank respectively. It should be noted that where

the superscripts are omitted in the notation it means that either of the two types of PageRanks is true.

Traditional (or normal) PageRank can be interpreted as the probability that a knowledgeable but mindless random walk hits a given Web page (node). This is because the random walk is aware of the addresses of all the Web pages but he chooses a next page to visit without considering any information that the page contains [17]. In other words, being at a particular node, the random walk jumps to any node with uniform probability if the node has no outgoing links, otherwise he chooses with uniform probability one of the outgoing links and goes to the selected node. Generally, the non-normalized traditional PageRank vector is defined below.

Definition 17.2 $\vec{R}_G^{(t)}$ for system S_G is defined as $\vec{R}_G^{(t)} = (\mathsf{I} - c\mathsf{A}_G^{\top})^{-1} n_G \vec{u}_G$, where I is an identity matrix of same size as A_G.

The PageRank $\vec{R}_G^{(t)}$ defined above is obtained from solving the system of linear equations arising from the eigenvalue problem $\vec{R}^{(1)} = \mathsf{M}_G \vec{R}^{(1)}$ with eigenvalue 1, where $\vec{R}^{(1)}$ is the normalized PageRank whose sum of all elements is 1 and M_G is the modified system matrix defined previously.

On the other hand, lazy PageRank is described by Lazy-random walk [7]. It differs from the traditional PageRank in that the walk on a graph has a 50% probability of staying or leaving each node. In other words, before choosing the next node to visit, the random walk first tosses a coin. If the head shows up he visits the next node, otherwise he stays in the very same node of the graph. The following proposition gives how the lazy PageRank is related to traditional PageRank.

Proposition 17.1 *Let $\vec{R}^{(t)}$ be a vector of the traditional PageRank of a graph G, then the lazy PageRank vector $\vec{R}^{(l)}$ is related to $\vec{R}^{(t)}$ by*

$$\vec{R}^{(t)}(\varepsilon, \vec{v}, \mathsf{A}_G) = \vec{R}^{(l)}\left(\frac{2\varepsilon}{1+\varepsilon}, \vec{v}, \mathsf{A}_G\right), \tag{17.1}$$

where $\varepsilon = 1 - c$ and $\vec{v}$ is a non-negative column vector whose elements sum up to 1.

Proof For traditional PageRank, the principal eigen-equation is given by

$$\vec{R}^{(t)} = (1-\varepsilon)\mathsf{A}_G^{\top}\vec{R}^{(t)} + \varepsilon\vec{v}.$$

In terms of lazy random walks, the equation is given by

$$\vec{R}^{(l)} = (1-\varepsilon)\frac{(\mathsf{I} + \mathsf{A}_G^{\top})}{2}\vec{R}^{(l)} + \varepsilon\vec{v}.$$

Multiplying both sides by 2 and simplifying we get

$$(1+\varepsilon)\vec{R}^{(l)} = (1-\varepsilon)\mathsf{A}_G^{\top}\vec{R}^{(l)} + 2\varepsilon\vec{v}.$$

Divide both sides by $(1+\varepsilon)$ to get

$$\vec{R}^{(l)} = \frac{1-\varepsilon}{1+\varepsilon}\mathsf{A}_G^{\top}\vec{R}^{(l)} + \frac{2\varepsilon}{1+\varepsilon}\vec{v}.$$

Hence (17.1) is proved. □

Proposition 17.1 reveals that the traditional PageRank $\vec{R}^{(t)}$ for ε, where ε is a fixed probability for the normal random walk to jump to any node in the graph, is the same as the lazy PageRank $\vec{R}^{(l)}$ for $\frac{2\varepsilon}{1+\varepsilon}$ using lazy random walk in the same graph.

The lazy PageRank of a graph is computed as follows.

Lemma 17.1 *The lazy PageRank $\vec{R}^{(l)}$ for a system S_G is given by*

$$\vec{R}^{(l)} = \left(\mathsf{I} - \beta\mathsf{A}_G^{\top}\right)^{-1}\alpha n_G\vec{u}_G, \tag{17.2}$$

where $\alpha = 2/(2-c)$, $\beta = c/(2-c)$ and I is an identity matrix of same size as A_G.

Proof By Proposition 17.1, we get

$$\vec{R}^{(l)} = \frac{c}{2-c}\mathsf{A}_G^{\top}\vec{R}^{(l)} + \frac{2(1-c)}{2-c}\vec{v} \Leftrightarrow \left(\mathsf{I} - \frac{c}{2-c}\mathsf{A}_G^{\top}\right)\vec{R}^{(l)} = \frac{2(1-c)}{2-c}\vec{v}$$

or

$$\vec{R}^{(l)} = \left(\mathsf{I} - \frac{c}{2-c}\mathsf{A}_G^{\top}\right)^{-1}\frac{2}{2-c}(1-c)\vec{v}$$

Letting $(1-c)\vec{v} = n_G\vec{u}_G$, we obtain the required result (17.2). □

It follows from Lemma 17.1 that in lazy-random walks, a random walk moves with probability $\beta \in (0,1)$ to a new vertex from the current vertex by traversing a random outgoing edge from the current vertex and stop the random walk with probability $1-\beta$ if the current vertex have no outgoing edge.

The non-normalized PageRank $\vec{R}_G$ also can be computed from a probabilistic viewpoint using random walks on a graph and the hitting probabilities of the said random walks. This is stated in the next definition.

Definition 17.3 Consider a random walk on a graph described by A_G, which is the adjacency matrix weighted such that the sum over every non-zero row is equal to one. In each step with probability $c \in (0,1)$, move to a new vertex from the current vertex by traversing a random outgoing edge from the current vertex with probability equal to the weight on the corresponding edge weight. With probability $1-c$ or if the current vertex have no outgoing edges, we stop the random walk. The PageRank $\vec{R}$ for a single vertex v_j can be written as

$$R_j = \left(\sum_{v_i \in V, v_i \neq v_j} w_i P_{ij} + w_j\right)\left(\sum_{k=0}^{\infty}(P_{jj})^k\right), \tag{17.3}$$

where P_{ij} is the probability to hit node v_j in a random walk starting in node v_i, before stopping of this random walk. This can be seen as the expected number of visits to v_j if we do multiple random walks, starting in every node once and weighting each of these random walks by $\vec{w}$ [9]. Note that we have both vertex and edge weights and that both the initial and final vertex are counted as being visited in such a random walk (but not counted extra).

Next, let us define graph-structures we will encounter in the section that follows.

Definition 17.4 A simple line is a graph with n_L nodes where node n_L links to node n_{L-1} which in turn links to node n_{L-2} all the way until node n_2 link to node n_1.

Definition 17.5 A complete graph is a group of nodes in which all nodes in the group links to all other nodes in the group.

The following well known lemma for blockwise inversion will be used in this article. A proof can be found, for example in Bernstein [3].

Lemma 17.2

$$\begin{bmatrix} \mathsf{B} & \mathsf{C} \\ \mathsf{D} & \mathsf{E} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathsf{B}-\mathsf{C}\mathsf{E}^{-1}\mathsf{D})^{-1} & -(\mathsf{B}-\mathsf{C}\mathsf{E}^{-1}\mathsf{D})^{-1}\mathsf{C}\mathsf{E}^{-1} \\ -\mathsf{E}^{-1}\mathsf{D}(\mathsf{B}-\mathsf{C}\mathsf{E}^{-1}\mathsf{D})^{-1} & \mathsf{E}^{-1}+\mathsf{E}^{-1}\mathsf{D}(\mathsf{B}-\mathsf{C}\mathsf{E}^{-1}\mathsf{D})^{-1}\mathsf{C}\mathsf{E}^{-1} \end{bmatrix}$$

where B, E *are square and* E, $(\mathsf{B}-\mathsf{C}\mathsf{E}^{-1}\mathsf{D})$ *are nonsingular.*

17.3 Changes in Traditional and Lazy PageRanks When Connecting the Simple Line with Multiple Outside Nodes

In this section, we present four graph-structures and associated PageRanks descriptions for both traditional PageRank and lazy PageRank. But first we present the following two results that are important in this section and in the paper as a whole.

Proposition 17.2 *The lazy PageRank of a node v_j with only outgoing edge(s) on a simple line is given by*

$$R_j^{(l)} = \frac{2}{2-c},$$

where $0 < c < 1$.

Proof Using Definition 17.3, the PageRank $\vec{R}^{(l)}$ for a single vertex v_j is given by

$$R_j^{(l)} = \left(\sum_{v_i \in V, v_i \neq v_j} w_i P_{ij} + w_j \right) \sum_{k=0}^{\infty} \left(P_{jj}\right)^k,$$

where P_{ij} is the probability to hit node v_j in a random walk starting from node v_i. Since v_j has no incoming edge, the weights $w_i = 0$ and $w_j = 1$. The probability of a random walk to stay at v_j in lazy random walk is $c/2$. Thus we have

$$R_j^{(l)} = \sum_{k=0}^{\infty} \left(\frac{2}{c}\right)^k = \frac{1}{1-c/2} = \frac{2}{2-c}.$$

□

Next we describe the lazy PageRank of any vertex v_j on the simple line as follows.

Lemma 17.3 *The lazy PageRank of a vertex v_j on the simple line can be expressed as*

$$R_j^{(l)} = \alpha \sum_{i=j}^{n_L} \beta^{i-j} = \alpha \left[\frac{1-\beta^{n_L-j+1}}{1-\beta}\right],$$

where $\alpha = 2/(2-c)$ and $\beta = c/(2-c)$.

Proof We prove this by induction. Letting $j = n_L$, the lazy PageRank of the last node v_L in the simple line follows directly from Proposition 17.2,

$$R_{n_L} = \alpha \sum_{i=n_L}^{n_L} \beta^{i-n_L} = \alpha = \frac{2}{2-c}.$$

Assume it is true for any node v_k, that is

$$R_k^{(l)} = \alpha \sum_{i=k}^{n_L} \beta^{i-k}, \tag{17.4}$$

we show that it's true for node v_{k-1} as well, which by induction, proves that it is generally true for all vertices in the simple line. Now, by Definition 17.3,

$$\begin{aligned} R_{k-1}^{(l)} &= \left(\sum_{\text{all} i, v_i \neq v_{k-1}} w_i P_{i,k-1} + w_{k-1}\right) \sum_{l=0}^{\infty} \left(P_{k-1,k-1}\right)^l \\ &= \left(\frac{1}{2} c R_k^{(l)} + 1\right) \alpha = \alpha + \beta R_k^{(l)}. \end{aligned}$$

Substituting (17.4) we obtain

$$R_{k-1}^{(l)} = \alpha + \alpha\beta \sum_{i=k}^{n_L} \beta^{i-k}$$

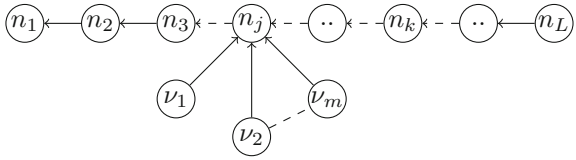

Fig. 17.1 A simple line with m outside vertices linked to one node on the line

$$= \alpha + \alpha \sum_{i=k}^{n_L} \beta^{i-k+1}$$
$$= \alpha \sum_{i=k-1}^{n_L} \beta^{i-(k-1)}.$$

□

17.3.1 Connecting the Simple Line with Multiple Links from Outside Nodes to One Node in the Line

Consider a simple line graph that has L vertices. Suppose vertex n_j, $j \in [1, L]$ is linked to m outside vertices as shown in Fig. 17.1. It can be seen that if $j = 1$, then the node is said to be an authority node.

Lemma 17.4 *The traditional PageRank of a node e_i belonging to the line in a system containing a simple line with m outside nodes linking to one node j in the line when using uniform weight vector $\vec{u}$ can be expressed as*

$$R_i^{(t)} = \sum_{k=0}^{n_L - i} c^k + b_{ij} = \frac{1 - c^{n_L - i + 1}}{1 - c} + b_{ij}$$
$$b_{ij} = \begin{cases} mc^{j-i+1}, & \text{if } i \leq j \\ 0, & \text{if } i > j \end{cases} \tag{17.5}$$

where $m \geq 1$ and n_L is the number of nodes in the line. The new nodes each have rank 1.

A proof of this lemma can be referred to our previous work in [5]. For the lazy PageRank, below is the theorem which describes similar behaviours for nodes in Fig. 17.1.

Theorem 17.1 *The lazy PageRank of a node e_i belonging to the line in a system containing a simple line with m outside nodes linking to one node j in the line when uniform weight vector $\vec{u}_G$ can be expressed as*

$$R_i^{(l)} = \alpha \sum_{k=i}^{n_L} \beta^{k-i} + b_{ij} = \alpha \left[\frac{1-\beta^{n_L-i+1}}{1-\beta} \right] + b_{ij}$$
$$b_{ij} = \begin{cases} m\alpha\beta^{j-i+1}, & \text{if } i \le j \\ 0, & \text{if } i > j \end{cases} \tag{17.6}$$

where $\alpha = 2/(2-c)$, $m \ge 1$ *and* n_L *is the number of nodes in the line. The new nodes each has rank* α.

Proof The proof of the lazy PageRank of the nodes on the line follows directly from Lemma 17.3. We only need to show that the m outside nodes linking to node e_j on the line add $b_{ij} = m\alpha\beta^{j-i+1}$ for $i \le j$.

Proving by induction, let $m = 1$. Then $b_{ij} = \alpha\beta^{j-i+1}$, and this is true for one outside node linking to e_j. Now assume that it holds for arbitrary number of nodes $m = k$,

$$b_{ij}(k) = \underbrace{\alpha\beta^{j-i+1} + \cdots + \alpha\beta^{j-i+1}}_{k \text{ times}} = k\alpha\beta^{j-i+1}.$$

For $m = k+1$,

$$\begin{aligned} b_{ij}(k+1) &= b_{ij}(k) + \alpha\beta^{j-i+1} = k\alpha\beta^{j-i+1} + \alpha\beta^{j-i+1} \\ &= (k+1)\alpha\beta^{j-i+1}. \end{aligned}$$

Lastly, the PageRank of the m outside nodes follows from Proposition 17.2. Thus each has PageRank $\alpha = 2/(2-c)$. □

17.3.2 Connecting a Simple Line with Multiple Links from Multiple Outside Nodes to the Line

Assume that the nodes $n_1, n_2, \cdots, n_{L-1}, n_L$ on the line are linked to outside nodes $m_1, m_2, \cdots, m_{L-1}, m_L$ respectively, where $m_j \ge 0$ (the number of outside nodes linked to node j on the simple line) as shown in the Fig. 17.2. The traditional PageRank for such general network for $m_j \ge 0$ is given in the next theorem. The proof can also be obtained in [5].

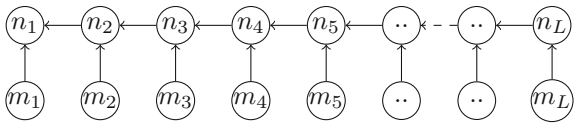

Fig. 17.2 A simple line with outside vertices linked to each vertex on the line

Theorem 17.2 *The PageRank of a node e_i belonging to the line in a system containing a simple line with multiple outside nodes, $m_1, m_2, \cdots, m_i, \ldots, m_L$ linking to every nodes $n_1, n_2, \cdots, n_i, \ldots, n_L$ in that order respectively in the line when using uniform weight vector $\vec{u}$ can be written as*

$$R_i^{(t)} = \frac{1 - c^{n_L - i + 1}}{1 - c} + \sum_{j=i}^{n_L} m_j c^{j-i+1}. \tag{17.7}$$

The outside nodes each have rank 1.

The corresponding theorem for lazy PageRank for the same graph-structure is given below.

Theorem 17.3 *The lazy PageRank $\vec{R}^{(l)}$ of a node e_i belonging to the line in a system containing a simple line with multiple outside nodes $m_1, m_2, \ldots, m_L$ linking to every nodes $n_1, n_2, \ldots, n_L$ in that order respectively can be expressed as*

$$R_i^{(l)} = \alpha \sum_{j=i}^{n_L} \beta^{j-i} + \sum_{j=i}^{n_L} m_j \alpha \beta^{j-i+1} \tag{17.8}$$

where $\alpha = 2/(2 - c)$. The outside nodes each has rank α.

Proof By induction, let's take only one node, say n_L. It is true to write

$$R_L^{(l)} = \alpha \sum_{j=L}^{L} \beta^{j-L} + \sum_{j=L}^{L} m_j \alpha \beta^{j-L+1} = \alpha + m_L \alpha \beta.$$

Suppose that it holds for any node k in the line, then

$$R_k^{(l)} = \alpha \sum_{j=k}^{n_L} \beta^{j-k} + \sum_{j=k}^{n_L} m_j \alpha \beta^{j-k+1}.$$

We need to show that it holds for node $k - 1$. That is

$$R_{k-1}^{(l)} = \alpha + \beta R_k^{(l)} + m_{k-1} \alpha \beta.$$

Substituting for $R_k^{(l)}$ we obtain

$$R_{k-1}^{(l)} = \alpha + \beta \left[\alpha \sum_{j=k}^{n_L} \beta^{j-k} + \sum_{j=k}^{n_L} m_j \alpha \beta^{j-k+1} \right] + m_{k-1} \alpha \beta$$

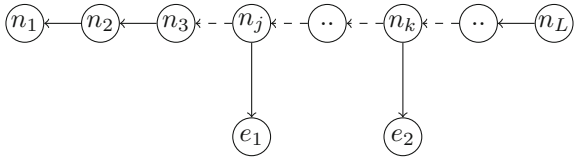

Fig. 17.3 A simple line with two links from the line to two outside nodes

$$= \alpha + \alpha\beta \sum_{j=k}^{n_L} \beta^{j-k} + \beta \sum_{j=k}^{n_L} m_j \alpha \beta^{j-k+1} + m_{k-1}\alpha\beta$$
$$= \alpha + \alpha \sum_{j=k}^{n_L} \beta^{j-(k-1)} + \sum_{j=k}^{n_L} m_j \alpha \beta^{j-(k-1)+1} + m_{k-1}\alpha\beta$$
$$= \alpha \sum_{j=k-1}^{n_L} \beta^{j-(k-1)} + \sum_{j=k-1}^{n_L} m_j \beta^{j-(k-1)+1}.$$

Finally, using Proposition 17.2, the outside nodes each has rank α. □

17.3.3 Connecting the Simple Line with Two Links from the Line to Two Outside Nodes

Suppose we consider a graph where two nodes in the line link to two outside nodes as in Fig. 17.3. A formulation (see [5] for the proof) below describes the traditional PageRank for the nodes.

Theorem 17.4 *The traditional PageRank $R_i^{(t)}$ of a node e_i belonging to the line in a system containing a simple line with two outside nodes, e_1 and e_2, whose links are from node j and k, $j < k$, respectively in the line when using uniform weight vector $\vec{u}$ can be expressed as*

$$R_i^{(t)} = \sum_{m=i}^{n_L} c^{m-i} = \frac{1 - c^{n_L - i + 1}}{1 - c}, \quad for \quad i \geq k, \tag{17.9}$$

$$R_i^{(t)} = \sum_{m=i+1}^{k} c^{m-i-1} + \frac{1}{2} \sum_{m=k}^{n_L} c^{m-i}$$
$$= \frac{2 - c^{k-i}\left(1 + c^{n_L - k + 1}\right)}{2(1-c)}, \quad for \quad j \leq i < k, \tag{17.10}$$

$$R_i^{(t)} = \sum_{m=i+1}^{j} c^{m-i-1} + \frac{1}{2}\sum_{m=j}^{k-1} c^{m-i} + \frac{1}{4}\sum_{m=k}^{n_L} c^{m-i}, \quad for \quad i < j$$
$$= \frac{4 - c^{k-i} - c^{j-i}\left(2 + c^{n_L-j+1}\right)}{4(1-c)}, \tag{17.11}$$

where n_L is the number of nodes in the line. The PageRank of the new nodes e_1 and e_2 are respectively,

$$R_{e_1}^{(t)} = 1 + \frac{1}{2}c\left(\frac{1-c^{k-j}}{1-c}\right) + \frac{1}{4}c^{k-j+1}\left(\frac{1-c^{n_L-k+1}}{1-c}\right) \tag{17.12}$$

and

$$R_{e_2}^{(t)} = 1 + \frac{1}{2}c\left(\frac{1-c^{n_L-k+1}}{1-c}\right). \tag{17.13}$$

On the other hand, the following describes the lazy PageRank for the same graph-structure.

Theorem 17.5 *The lazy PageRank $R_i^{(l)}$ of a node e_i belonging to the line in a system containing a simple line with two outside nodes, e_1 and e_2, whose links are from node j and k, $j < k$, respectively in the line when using uniform weight vector $\vec{u}$ can be expressed as*

$$R_i^{(l)} = \alpha\sum_{m=i}^{n_L}\beta^{m-i} = \alpha\left[\frac{1-\beta^{n_L-i+1}}{1-\beta}\right], \quad for \quad i \geq k, \tag{17.14}$$

$$R_i^{(l)} = \alpha\sum_{m=i+1}^{k}\beta^{m-i-1} + \frac{\alpha}{2}\sum_{m=k}^{n_L}\beta^{m-i}, \quad for \quad j \leq i < k, \tag{17.15}$$

$$R_i^{(l)} = \alpha\sum_{m=i+1}^{j}\beta^{m-i-1} + \frac{\alpha}{2}\sum_{m=j}^{k-1}\beta^{m-i} + \frac{\alpha}{4}\sum_{m=k}^{n_L}\beta^{m-i}, \quad for \quad i < j, \tag{17.16}$$

where n_L is the number of nodes in the line. The lazy PageRank of the new nodes e_1 and e_2 are respectively,

$$R_{e_1}^{(l)} = \alpha + \frac{\alpha}{2}\sum_{m=j+1}^{k}\beta^{m-j} + \frac{\alpha}{4}\sum_{m=k}^{n_L}\beta^{m-j+1} \tag{17.17}$$

and

$$R_{e_2}^{(l)} = \alpha + \frac{\alpha}{2}\sum_{m=k}^{n_L}\beta^{m-k+1}. \tag{17.18}$$

Proof The lazy PageRank of a node e_i, for $i \geq k$ follows from Lemma 17.3. Hence, (17.14) is proved similarly to the said lemma.

For $j \leq i < k$, the expected number of hits to node e_i in a lazy random walk starting from all nodes between j and k on the simple line is given by the first term of (17.15), i.e

$$\alpha \sum_{m=i+1}^{k} \beta^{m-i-1}.$$

The second term of the (17.15) represents the expected number of hits to node e_i starting from all nodes $k, k+1, \ldots, n_{L-1}, n_L$, i.e

$$\frac{\alpha}{2} \sum_{m=k}^{n_L} \beta^{m-i},$$

where the $1/2$ before the summation symbol accounts for the number of outgoing edges at node k. Hence, by Definition 17.3, the PageRank $R_i^{(l)}$ of node e_i is the expected number of hits to e_i if we do multiple lazy random walks starting in every node on the line for which $j \leq i < k$ is true. This is then given by adding the two terms above which gives (17.15).

Similarly, the lazy PageRank of node e_i, for $i < j$ in the simple line is the sum of all the expected number of hits to e_i in lazy random walk starting from all nodes $i+1, i+2, \ldots, n_L$. That is

$$R_i^{(l)} = \alpha \sum_{m=i+1}^{j} \beta^{m-i-1} + \frac{\alpha}{2} \sum_{m=j}^{k-1} \beta^{m-i} + \frac{\alpha}{4} \sum_{m=k}^{n_L} \beta^{m-i}, \quad \text{for} \quad i < j,$$

where the first term is the expected number of hits to e_i starting from all nodes between i and j, the second is the expected number of hits to e_i starting from nodes j to $k-1$ and the third term is the expected number of hits to e_i starting from node k to n_L in the line. The quarter multiplied to the third term caters for the number of outgoing edges at j and k positions in the line.

Finally, to get the PageRanks $R_{e_1}^{(l)}$ and $R_{e_2}^{(l)}$ we consider all lazy random walks to hit e_1 and e_2, respectively. Using Definition 17.3, we have $R_{e_1}^{(l)} = \left(1 + \frac{1}{4} c R_j^{(l)}\right) \alpha$, where $R_j^{(l)} = \alpha \sum_{m=j+1}^{k} \beta^{m-j-1} + \frac{\alpha}{2} \sum_{m=k}^{n_L} \beta^{m-j}$. Upon substitution and simplification we get the desired result

$$R_{e_1}^{(l)} = \alpha + \frac{\alpha}{2} \sum_{m=j+1}^{k} \beta^{m-j} + \frac{\alpha}{4} \sum_{m=k}^{n_L} \beta^{m-j+1}.$$

Similarly, the PageRank $R_{e_2}^{(l)}$ is given by $R_{e_2}^{(l)} = \left(1 + \frac{1}{4} c R_k^{(l)}\right) \alpha$, where $R_k^{(l)} = \alpha \sum_{m=k}^{n_L} \beta^{m-k}$. If we substitute and simplify we get $R_{e_2}^{(l)} = \alpha + \frac{\alpha}{2} \sum_{m=k}^{n_L} \beta^{m-k+1}$. □

17.4 Changes in Traditional and Lazy PageRanks When Connecting the Simple Line with Two Links from the Line to Two Complete Graphs

Consider a network S in which a simple line S_L is connected to two complete graphs S_{G_1} and S_{G_2} such that $S = S_L \cup S_{G_1} \cup S_{G_2}$. We assume that the subgraphs S_{G_1} and S_{G_2} each is linked to nodes j and k respectively as shown in Fig. 17.4. Then we will describe the two variants of PageRank for the nodes in the given network. But first we give two necessary previous results regarding traditional PageRank for a complete graph by Engström and Silvestrov [9].

Lemma 17.5 *The diagonal element a_d of the inverse $(\mathsf{I} - c\mathsf{A}_G^\top)^{-1}$ of the complete graph with n nodes is*

$$a_d = \frac{(n-1) - c(n-2)}{(n-1) - c(n-2) - c^2}. \tag{17.19}$$

The non diagonal elements a_{ij} can be written as

$$a_{ij} = \frac{c}{(n-1) - c(n-2) - c^2}. \tag{17.20}$$

A proof of this result can be referred to [9]. Similarly, for a lazy PageRank, the diagonal element a_d of the inverse $(\mathsf{I} - \beta\mathsf{A}_G^\top)^{-1}$ of the complete graph with n nodes is

$$a_d = \frac{(n-1) - \beta(n-2)}{(n-1) - \beta(n-2) - \beta^2} \tag{17.21}$$

and

$$a_{ij} = \frac{\beta}{(n-1) - \beta(n-2) - \beta^2} \tag{17.22}$$

for non diagonal elements a_{ij}, where $\beta = c/(2-c)$.

Theorem 17.6 *Given a complete graph with $n > 1$ nodes, PageRank $\vec{R}^{(t)}$ before normalization can be written as*

$$R_i^{(t)} = \frac{1}{1-c}. \tag{17.23}$$

The proof of this result can also be referred to [9]. The following theorem describes the relationship between traditional and lazy PageRanks for a complete graph.

Theorem 17.7 *Given a complete graph with $n > 1$ nodes, the PageRanks $\vec{R}^{(t)}$ and $\vec{R}^{(l)}$ before normalization are the same and equals to*

$$R_i^{(t)} = R_i^{(l)} = \frac{1}{1-c}. \tag{17.24}$$

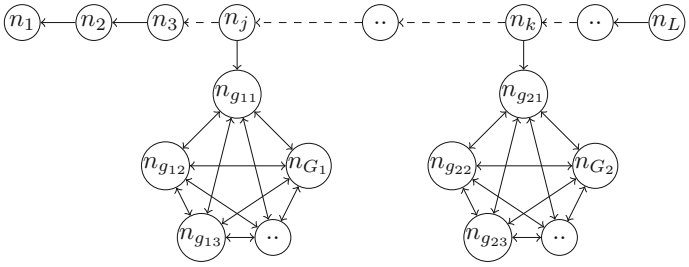

Fig. 17.4 A simple line with two links from the line to two complete graphs

Proof Using Definition 17.2 where the system S_G is the complete graph, the traditional PageRank $\vec{R}^{(t)} = (\mathsf{I} - c\mathsf{A}_G^\top)^{-1} n\vec{u}_G$ is a vector whose elements are identical. From Theorem 17.6, the ith element is expressed as $R_i^{(t)} = \frac{1}{1-c}$.

Similarly, using Lemma 17.1, the lazy PageRank of a complete graph with $n > 1$ nodes is given by $\vec{R}^{(l)} = \alpha \left(\mathsf{I} - \beta\mathsf{A}_G^\top\right)^{-1} n\vec{u}_G$. Analogously, $\left(\mathsf{I} - \beta\mathsf{A}_G^\top\right)^{-1} n\vec{u}_G$ is the vector whose elements are equal to $1/(1-\beta)$. Hence $R_i^{(l)} = \frac{\alpha}{1-\beta} = \frac{1}{1-c}$ since $\alpha = 2/(2-c)$ and $\beta = c/(2-c)$. □

Now, we describe the PageRanks of the system given in Fig. 17.4 as follows.

Theorem 17.8 *Let S be a system made up of three systems: a simple line S_L with n_L nodes, two complete graphs, S_{G_1} and S_{G_2}, with n_{G_1} and n_{G_2} nodes, respectively. We add two links from nodes j and k, $j < k$ in the line to nodes g_j and g_k in the first and second complete graph, respectively. Assuming uniform weight vector $\vec{u}$, we get the PageRank $R_{L,i}^{(t)}$, where $S_G = S_{G_1} \cup S_{G_2}$, for the nodes in the line after the new links, $R_{G_1,i}^{(t)}$ for the nodes in the first complete graph S_{G_1} and $R_{G_2,i}^{(t)}$ for the nodes in the second complete graph S_{G_2} as:*

$$R_{L,i}^{(t)} = \sum_{m=i}^{n_L} c^{m-i} = \frac{1 - c^{n_L - i + 1}}{1-c}, \quad \textit{for} \quad i \geq k, \tag{17.25}$$

$$\begin{aligned} R_{L,i}^{(t)} &= \sum_{m=i+1}^{k} c^{m-i-1} + \frac{1}{2}\sum_{m=k}^{n_L} c^{m-i}, \quad \textit{for} \quad j \leq i < k \\ &= \frac{2 - c^{k-i}\left(1 + c^{n_L - k + 1}\right)}{2(1-c)}, \quad \textit{for} \quad j \leq i < k, \end{aligned} \tag{17.26}$$

$$\begin{aligned} R_{L,i}^{(t)} &= \sum_{m=i+1}^{j} c^{m-i-1} + \frac{1}{2}\sum_{m=j}^{k-1} c^{m-i} + \frac{1}{4}\sum_{m=k}^{n_L} c^{m-i}, \quad \textit{for} \quad i < j \\ &= \frac{4 - c^{k-i} - c^{j-i}\left(2 + c^{n_L - j + 1}\right)}{4(1-c)}, \end{aligned} \tag{17.27}$$

where n_L is the number of nodes in the line. The traditional PageRank of the nodes in the complete graphs are given by

$$R^{(t)}_{G_2,g_k} = \frac{c}{2}\left(\frac{1-c^{n_L-k+1}}{1-c}\right)\left(\frac{(n_{G_2}-1)-c(n_{G_2}-2)}{(n_{G_2}-1)-c(n_{G_2}-2)-c^2}\right) + \frac{1}{1-c} \tag{17.28}$$

$$R^{(t)}_{G_2,i} = \left(\frac{c^2(1-c^{n_L-k+1})}{2(1-c)}\right)\left(\frac{1}{(n_{G_2}-1)-c(n_{G_2}-2)-c^2}\right) + \frac{1}{1-c} \tag{17.29}$$

$$R^{(t)}_{G_1,g_j} = \left[\frac{2c-c^{k-j+1}-c^{n_L-j+2}}{4(1-c)}\right]\left[\frac{(n_{G_1}-1)-c(n_{G_1}-2)}{(n_{G_1}-1)-c(n_{G_1}-2)-c^2}\right] + \frac{1}{1-c} \tag{17.30}$$

$$R^{(t)}_{G_1,i} = \left[\frac{2c^2-c^{k-j+2}-c^{n_L-j+3}}{4(1-c)}\right]\left[\frac{1}{(n_{G_1}-1)-c(n_{G_1}-2)-c^2}\right] + \frac{1}{1-c} \tag{17.31}$$

where $R^{(t)}_{G_1,g_j}$ is the traditional PageRank for the node in the complete graph S_{G_1} linked by the line and $R^{(t)}_{G_1,i}$ is the traditional PageRank of the other nodes in S_{G_1}. Similarly, $R^{(t)}_{G_2,g_k}$ is the traditional PageRank for the node in the complete graph S_{G_2} linked by the line and $R^{(t)}_{G_2,i}$ is the traditional PageRank of the other nodes in S_{G_2}.

For the proof of this theorem, one is advised to refer to our previous work [5]. Correspondingly, the next theorem describes the lazy PageRank for the same network structure.

Theorem 17.9 *Let S be a system made up of three systems: a simple line S_L with n_L nodes, two complete graphs, S_{G_1} and S_{G_2}, with n_{G_1} and n_{G_2} nodes, respectively. We add two links from nodes j and k, $j < k$ in the line, to nodes g_j and g_k in S_{G_1} and S_{G_2}, respectively. Assuming uniform weight vector $\vec{u}$, we get the following lazy PageRanks: $R^{(l)}_{L,i}$ for the nodes in the line after the new links, $R^{(l)}_{G_1,i}$ for the nodes in the first complete graph S_{G_1} and $R^{(l)}_{G_2,i}$ for the nodes in the second complete graph S_{G_2} as:*

$$R^{(l)}_{L,i} = \alpha\sum_{m=i}^{n_L}\beta^{m-i} = \alpha\left[\frac{1-\beta^{n_L-i+1}}{1-\beta}\right], \quad for \quad i \geq k, \tag{17.32}$$

$$R^{(l)}_{L,i} = \alpha\sum_{m=i+1}^{k}\beta^{m-i-1} + \frac{\alpha}{2}\sum_{m=k}^{n_L}\beta^{m-i}, \quad for \quad j \leq i < k, \tag{17.33}$$

$$R^{(l)}_{L,i} = \alpha\sum_{m=i+1}^{j}\beta^{m-i-1} + \frac{\alpha}{2}\sum_{m=j}^{k-1}\beta^{m-i} + \frac{\alpha}{4}\sum_{m=k}^{n_L}\beta^{m-i}, \quad for \quad i < j, \tag{17.34}$$

where n_L is the number of nodes in the line. The lazy PageRank of the nodes in the complete graphs are given by

$$R^{(l)}_{G_2,g_k} = \frac{\alpha\beta}{2}\left[\frac{1-\beta^{n_L-k+1}}{1-\beta}\right] \cdot \left[\frac{(n_{G_2}-1)-\beta(n_{G_2}-2)}{(n_{G_2}-1)-\beta(n_{G_2}-2)-\beta^2}\right] + \frac{1}{1-c} \tag{17.35}$$

$$R^{(l)}_{G_2,i} = \frac{\alpha\beta^2}{2}\left[\frac{(1-\beta^{n_L-k+1})}{(1-\beta)}\right] \cdot \left[\frac{1}{(n_{G_2}-1)-\beta(n_{G_2}-2)-\beta^2}\right] + \frac{1}{1-c} \tag{17.36}$$

$$R^{(l)}_{G_1,g_j} = \frac{\alpha\beta}{2}\left[\sum_{m=j+1}^{k}\beta^{m-j-1} + \frac{1}{2}\sum_{m=k}^{n_L}\beta^{m-j}\right] \cdot \left[\frac{(n_{G_1}-1)-\beta(n_{G_1}-2)}{(n_{G_1}-1)-\beta(n_{G_1}-2)-\beta^2}\right] + \frac{1}{1-c} \tag{17.37}$$

$$R^{(l)}_{G_1,i} = \frac{\alpha\beta^2}{2}\left[\sum_{m=j+1}^{k}\beta^{m-j-1} + \frac{1}{2}\sum_{m=k}^{n_L}\beta^{m-j}\right] \cdot \left[\frac{1}{(n_{G_1}-1)-\beta(n_{G_1}-2)-\beta^2}\right] + \frac{1}{1-c} \tag{17.38}$$

where $R^{(l)}_{G_1,g_j}$ is the lazy PageRank for the node in the complete graph S_{G_1} linked by the line and $R^{(l)}_{G_1,i}$ is the lazy PageRank of the other nodes in S_{G_1}. Similarly, $R^{(l)}_{G_2,g_k}$ is the lazy PageRank for the node in the complete graph S_{G_2} linked by the line and $R^{(l)}_{G_2,i}$ is the lazy PageRank of the other nodes in S_{G_2}. Note that $\alpha = 2/(2-c)$ and $\beta = \alpha c/2$.

Proof Let use the blockwise inversion Lemma 17.2 to find the inverse of $(\mathsf{I} - \beta\mathsf{A}_G^\top)$, expressed as $(\mathsf{I} - \beta\mathsf{A}_G^\top) = \begin{bmatrix}\mathsf{B} & \mathsf{C}\\ \mathsf{D} & \mathsf{E}\end{bmatrix}$, where the sizes of the matrices are given as $\mathsf{B} : n_L \times n_L$, $\mathsf{C} : n_L \times n_G$, $\mathsf{D} : n_G \times n_L$ and $\mathsf{B} : n_G \times n_G$, for $n_G = n_{G_1} + n_{G_2}$. We write the inverse as $(\mathsf{I} - \beta\mathsf{A}_G^\top)^{-1} = \begin{bmatrix}\mathsf{B}^{inv} & \mathsf{C}^{inv}\\ \mathsf{D}^{inv} & \mathsf{E}^{inv}\end{bmatrix}$, where by Lemma 17.2, $\mathsf{B}^{inv} = (\mathsf{B} - \mathsf{C}\mathsf{E}^{-1}\mathsf{D})^{-1} = \mathsf{B}^{-1}$, $\mathsf{C}^{inv} = -\mathsf{B}^{-1}\mathsf{C}\mathsf{E}^{-1} = \mathsf{O}$, since C consists of zero entries, $\mathsf{D}^{inv} = -\mathsf{E}^{-1}\mathsf{D}\mathsf{B}^{-1}$ and $\mathsf{E}^{inv} = \mathsf{E}^{-1}$.

We know that matrices B and E are invertible. Therefore

$$\mathsf{B}^{-1} = \begin{bmatrix} 1 & \beta & \beta^2 & \cdots & \frac{\beta^{j-1}}{2} & \cdots & \frac{\beta^{k-2}}{2} & \frac{\beta^{k-1}}{4} & \cdots & \frac{\beta^{n_L-1}}{4} \\ 0 & 1 & \beta & \cdots & \frac{\beta^{j-2}}{2} & \cdots & \frac{\beta^{k-3}}{2} & \frac{\beta^{k-2}}{4} & \cdots & \frac{\beta^{n_L-2}}{4} \\ \vdots & \ddots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ & & & & 1 & \beta & \cdots & \frac{\beta^{k-j}}{2} & \cdots & \frac{\beta^{n_L-j}}{2} \\ & & & & & \ddots & & \vdots & & \vdots \\ & & & & & & & 1 & \cdots & \beta^{n_L-k} \\ & & & & & & & & & \vdots \\ \vdots & & & & & & & & \ddots & \beta \\ 0 & 0 & \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix}$$

and $\mathsf{E}^{-1} = \begin{bmatrix} \mathsf{E}_1^{inv} & \mathsf{O} \\ \mathsf{O} & \mathsf{E}_2^{inv} \end{bmatrix}$, where E_1 and E_2 are weighted adjacency matrices corresponding to complete graphs S_{G_1} and S_{G_2}, respectively. Suppose E_1 of size $n_{G_1} \times n_{G_1}$ takes the form

$$\mathsf{E}_1 = \begin{bmatrix} 1 & a_1 & a_1 & \cdots & a_1 \\ a_1 & 1 & a_1 & \cdots & a_1 \\ a_1 & a_1 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & a_1 \\ a_1 & a_1 & \cdots & a_1 & 1 \end{bmatrix},$$

where $a_1 = -\beta/(n_{G_1} - 1)$. Then from (17.21) and (17.22),

$$\mathsf{E}_1^{inv} = \mathsf{E}_1^{-1} = \begin{bmatrix} a_d & k_1 & k_1 & \cdots & k_1 \\ k_1 & a_d & k_1 & \cdots & k_1 \\ k_1 & k_1 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & k_1 \\ k_1 & k_1 & \cdots & a_1 & a_d \end{bmatrix},$$

where $a_d = 1 - (n_{G_1} - 1)a_1k_1$ and $k_1 = \frac{\beta}{(n_{G_1}-1)-\beta(n_{G_1}-2)-\beta^2}$. Substituting for a_1 and k_1 in a_d, we get

$$a_d = \frac{(n_{G_1} - 1) - (n_{G_1} - 2)\beta}{(n_{G_1} - 1) - \beta(n_{G_1} - 2) - \beta^2}.$$

In a similar way, we obtain E_2^{-1} as a square matrix of size $n_{G_2} \times n_{G_2}$ whose diagonal element is b_d and off diagonal element is k_2, where $b_d = \frac{(n_{G_2}-1)-(n_{G_2}-2)\beta}{(n_{G_2}-1)-\beta(n_{G_2}-2)-\beta^2}$ and $k_2 = \frac{\beta}{(n_{G_2}-1)-\beta(n_{G_2}-2)-\beta^2}$.

Now, we compute D^{inv} as $\mathsf{D}^{inv} = -\mathsf{E}^{-1}\mathsf{D}\mathsf{B}^{-1}$. Since $\mathsf{D}\mathsf{B}^{-1}$ can be expressed

$$\mathsf{DB}^{-1} = -\frac{1}{2}\beta \begin{bmatrix} 0 \cdots 0 & 1 & \beta/2 & \cdots & \beta^{n_L-j} \\ 0 \quad \cdots \quad 0 & & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 \quad \cdots \quad 0 & 1 & \beta & \cdots & \beta^{n_L-k} \\ \vdots & & 0 & \cdots & 0 \\ & & \vdots & & \vdots \\ 0 \quad \cdots \quad 0 & & 0 & \cdots & 0 \end{bmatrix},$$

then, the inverse of D is given by

$$\mathsf{D}^{-1} = \begin{bmatrix} 0 \cdots 0 & \frac{1}{2}k_1\beta & \frac{1}{4}a_d\beta^2 & \cdots & \frac{1}{4}a_d\beta^{n_L-k} & \cdots & \frac{1}{4}a_d\beta^{n_L-j} \\ 0 \cdots 0 & \frac{1}{2}k_1\beta & \frac{1}{4}k_1\beta^2 & \cdots & \frac{1}{4}k_1\beta^{n_L-j} & \cdots & \frac{1}{4}k_1\beta^{n_L-j} \\ \vdots \quad \vdots & & \vdots & & & & \vdots \\ 0 \cdots 0 & \frac{1}{2}k_1\beta & \frac{1}{4}k_1\beta^2 & \cdots & \frac{1}{4}k_1\beta^{n_L-k} & \cdots & \frac{1}{4}k_1\beta^{n_L-j} \\ 0 \cdots 0 & 0 & 0 & \cdots & \frac{1}{2}b_d\beta & \cdots & \frac{1}{2}b_d\beta^{n_L-k} \\ 0 \cdots 0 & 0 & 0 & \cdots & \frac{1}{2}k_2\beta & \cdots & \frac{1}{2}k_2\beta^{n_L-k} \\ \vdots & & & & \vdots & & \vdots \\ 0 \cdots 0 & 0 & 0 & \cdots & \frac{1}{2}k_2\beta & \cdots & \frac{1}{2}k_2\beta^{n_L-k} \end{bmatrix},$$

where k_1, k_2, a_d and b_d are as defined above. Since

$$\mathsf{E}^{-1} = \begin{bmatrix} \mathsf{E}_1^{-1} & \mathsf{O} \\ \mathsf{O} & \mathsf{E}_2^{-1} \end{bmatrix},$$

then $(\mathsf{I} - \beta\mathsf{A}_G^\top)^{-1}$ is known. One need only to sum across the rows and multiply by α to obtain the lazy PageRank of each node. This ends the proof. □

17.4.1 A Comparison of Traditional PageRank and Lazy PageRank for the Line Connected with Complete Graphs

In this subsection, we present the ranking behaviour of both non-normalized traditional PageRank and lazy PageRank for the simple line connected with two complete graphs as in Fig. 17.4. We do this by looking at the ranks as a function of c. We consider arbitrary characteristic of the graph, i.e., a graph with $n_{G_1} = n_{G_2} = 5$, $n_L = 7$, $j = 3$ and $k = 5$.

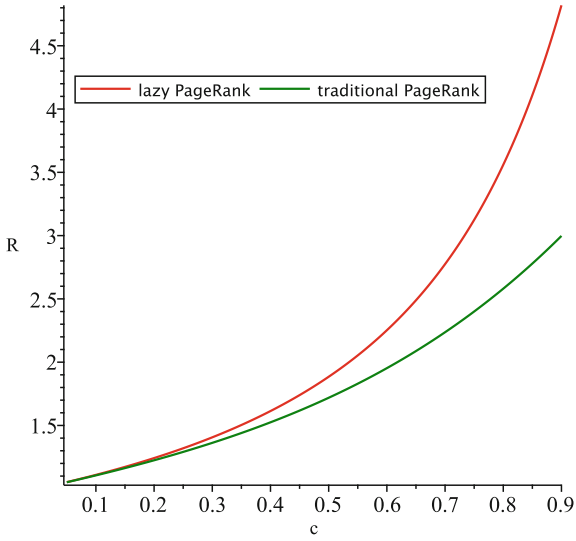

Fig. 17.5 $\vec{R}$ of the node on the line linking to one of the complete graphs as a function of c

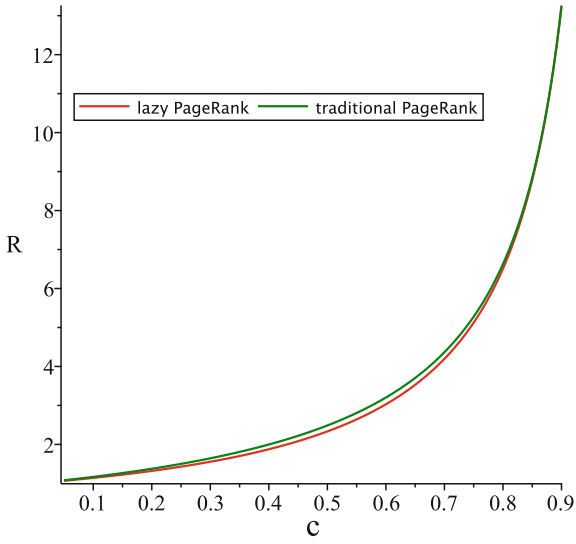

Fig. 17.6 $\vec{R}$ of a node in the complete graph linked from the line as a function of c

Specifically, we are interested in the PageRank on the node n_k on the simple line, $n_{g_{21}}$ and any other node in the complete graph, see Fig. 17.4 for further details. In such cases we wish to provide a comparison for the two types of PageRank.

Figure 17.5 indicates the variation of the two variants of PageRank for node n_k. Results reveals that as c approaches 0.99, a lazy random walk has higher expected number of visits than traditional random walk if multiple random walks are performed, starting in every node once. Figure 17.6 presents a plot of $R_{n_{g_{21}}}$ against c. The findings show that at about $c = 0.80$, both traditional and lazy PageRanks seem

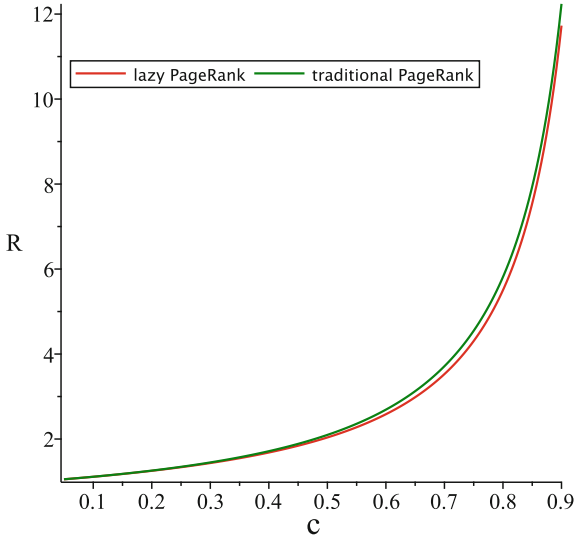

Fig. 17.7 $\vec{R}$ of the other nodes in the complete graph not linked directly from the line as a function of c

to be the same rank. This suggests that if one takes into account the lazy random walk of surfers in a complete graph, such consideration might not yield any difference as compared to traditional random walk, particularly when c is small (≤ 0.25) and higher (≥ 0.80), whereas, in Fig. 17.7, no substantial difference is observed for $c \leq 0.5$. In general, for complete graphs traditional and lazy PageRanks have quite slight difference in numerical rank values (Figs. 17.6 and 17.7).

17.5 Conclusions

In this article, we have derived explicit formulae and compare the two variants of PageRank. We started with the derivation of the formulae using non-normalized PageRank technique for both traditional and lazy PageRanks for a line of nodes connected to complete graphs. Thereafter, we tackled how the variants of PageRank differ as a function of damping factor, c. To achieve this, we compared ranks at three locations on the graph, that is, a node located on a line, a node on a complete graph linked to a line and a node solely contained in a complete graph. We have observed that a lazy random walk spend longer time on a node in a line than traditional random walk. Whereas, for any node which is part of the complete graph, the difference between the two random walks is insignificant.

Acknowledgements This research was supported by the Swedish International Development Cooperation Agency (Sida), International Science Programme (ISP) in Mathematical Sciences (IPMS), Sida Bilateral Research Program (Makerere University and University of Dar-es-Salaam). We are also grateful to the research environment Mathematics and Applied Mathematics (MAM), Division

of Applied Mathematics, Mälardålen University for providing an excellent and inspiring environment for research education and research. The authors are grateful to Dmitrii Silvestrov for useful comments and discussions.

References

1. Anderson, F., Silvestrov, S.: The mathematics of internet search engines. Acta Appl. Math. **104**, 211–242 (2008)
2. Battiston, S., Puliga, M., Kaushik, R., Tasca, P., Calderelli, G.: DeptRank: Too central to fail? Financial networks, the fed and systemic risk. Technical Report (2012)
3. Bernstein, D.: Matrix Mathematics. Princeton Unversity Press, NY (2005)
4. Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences. Society for industrial and applied mathematics (1994)
5. Biganda, P.S., Abola, B., Engström, C., Silvestrov, S.: PageRank, connecting a line of nodes with multiple complete graphs. In: Proceedings of the 17th Conference of the ASMDA International Society Applied Stochastic Models and Data Analysis (ASMDA 2017), pp. 125–138 (2017)
6. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems **30**((1–7)), 107–117 (1998). Proceedings of the Seventh International World Wide Web Conference
7. Chung, F., Zhao, W.: PageRank and random walks on graphs (2008). http://www.math.ucsd.edu/~fan/wp/lov.pdf
8. Dhyani, D., Bhowmick, S.S., Ng, W.K.: Deriving and verifying statistical distribution of a hyperlink-based web page quality metric. Data Knowl. Eng. **46**(3), 291–315 (2003)
9. Engström, C., Silvestrov, S.: PageRank, a look at small changes in a line of nodes and the complete graph. In: Silvestrov S., Rančić M. (eds) Engineering Mathematics II. Springer Proceedings in Mathematics and Statistics, vol 179. Springer, Cham, 223-247 (2016)
10. Engström, C., Silvestrov, S.: PageRank, connecting a line of nodes with a complete graph. In: Silvestrov, S., Rančić, M. (eds.) Engineering Mathematics II. Springer proceedings in mathematics and statistics, vol. 179, pp. 249–274. Springer, Cham (2016)
11. Engström, C., Silvestrov, S.: A componentwise PageRank algorithm. In: Bozeman, J.R., Oliveira, T., Skiadas, C.H. (eds.) Stochastic and Data Analysis Methods and Applications in Statistics and Demography, pp. 375–388 (2016). ASMDA 2015 Proceedings: 16th Applied Stochastic Models and Data Analysis International Conference with 4th Demographics 2015 Workshop.,ISAST: International Society for the Advancement of Science and Technology
12. Haveliwala, T., Kamvar, S.: The second eigenvalue of the google matrix. Technical Report 2003–20, Stanford InfoLab (2003)
13. Ishii, H., Tempo, R., Bai, E.W., Dabbene, F.: Distributed randomized pagerank computation based on web aggregation. In: Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC, pp. 3026–3031 (2009)
14. Kamvar, S.D., Haveliwala, T.: The Condition Number of the PageRank Problem. Technical Report, Stanford InfoLab (2003)
15. Kamvar, S., Haveliwala, T., Golub, G.: Adaptive methods for the computation of pagerank. Linear Algebra Appl. **386**, 51–65 (2004)
16. Kamvar, S.D., Schlosser, M.T., Garcia-Molina, H.: The eigentrust algorithm for reputation management in p2p networks. In: Proceedings of the 12th International Conference On World Wide Web, WWW '03, pp. 640-651 (2003)
17. Klopotek, M.A., Wierzchon, S.T., Ciesielski, K., Czerski, D., Draminski, M.: Lazy walks versus walks with backstep: flavor of PageRank. In: Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Vol. 01, pp. 262–265 (2014)
18. Norris, J.R.: Markov Chains. Cambridge University Press, New York (2009)

Chapter 18
Continuous Approximations of Discrete Choice Models Using Point Process Theory

Hannes Malmberg and Ola Hössjer

Abstract We analyze continuous approximations of discrete choice models with a large number of options. We start with a discrete choice model where agents choose between different options, and where each option is defined by a characteristic vector and a utility level. For each option, the characteristic vector and the utility level are random and jointly dependent. We analyze the optimal choice, which we define as the characteristic vector of the option with the highest utility level. This optimal choice is a random variable. The continuous approximation of the discrete choice model is the distributional limit of this random variable as the number of offers tends to infinity. We use point process theory and extreme value theory to derive an analytic expression for the continuous approximation, and show that this can be done for a range of distributional assumptions. We illustrate the theory by applying it to commuting data. We also extend the initial results by showing how the theory works when characteristics belong to an infinite-dimensional space, and by proposing a setup which allows us to further relax our distributional assumptions.

Keywords Discrete choice · Random utility · Extreme value theory · Random fields · Point processes · Concomitant of order statistics

H. Malmberg (✉)
Stanford Institute for Economic Policy Research, 366 Galvez Street, Stanford, CA 94305-6015, USA
email: malmberg@stanford.edu

H. Malmberg
Department of Economics, University of Minnesota, 4-101 Hanson Hall 1925 4th Street S, Minneapolis, MN 55455, USA
email: pmalmber@umn.edu

O. Hössjer
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
e-mail: ola@math.su.se

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_18

18.1 Introduction

There is a long tradition in economics to use random utility theory to study discrete choices such as the choice of mode of transportation. Early contributions are Luce [12] and Mcfadden [15]. Over time, random utility theory has been extended to encompass more functional forms, distributional assumptions, and applications (Ben-Akiva and Lerman [2], Anderson et al. [1], Train [20]). The theory posits that agents maximize utility, but that utility is random from the econometrician's point of view. Utility is expressed as a random variable

$$U_i = f(X_i) + \varepsilon_i \quad i = 1, \ldots, n,$$

where U_i is the utility of option i, X_i are random variables that describe the characteristics of option i, $f(X_i)$ is the deterministic component of utility, and ε_i are independently and identically distributed random variables. The agent chooses the option with the highest utility.

Insofar each option has distinct characteristics, we can equivalently view this as a choice over the characteristics X_i. We write $X_{[n:n]}$ for the X_i corresponding to the largest U_i. This is a random variable taking values in the set $\{X_1, \ldots, X_n\} \subseteq \Omega$, where Ω is a general characteristics space.

We are interested in a continuous approximation to the discrete choice problem when the number of options is large, and we define the approximation as the distributional limit of the law of $X_{[n:n]}$ as $n \to \infty$. A continuous approximation takes an offer distribution density Λ, a deterministic utility component $f(\cdot)$, and the distribution of the random utility component ε_i, as inputs. The output is a probability distribution of choices over Ω.

The theory is relevant in situations where agents face discrete choices and a large number of options. For example, the choice of residential location in a city is a discrete choice as agents only buy one residence. This makes a random utility approach natural. On the other hand, the number of potential residential locations is large. In this case, it can be useful to approximate the discrete choice process with a continuous probability distribution over space.

We approach the problem by interpreting the collection of characteristics-utility pairs $(X_1, U_1), \ldots, (X_n, U_n)$ as the realizations of a point process ξ_n on the Cartesian product $\Omega \times \mathbb{R}$ of the characteristics space and the utility space. With this interpretation, the best choice $X_{[n:n]}$ is a function of ξ_n. More details on point process theory can be found, for instance, in Cox and Isham [4] and Jacobsen [9]. More specifically, we can build on the results in point process theory presented in Resnick [17] to derive sharp results on the limiting behavior of $X_{[n:n]}$. In particular, we show that a monotone transformation of the underlying point process ξ_n converges to a Poisson process on $\Omega \times \mathbb{R}$ and we derive the limiting behavior of $X_{[n:n]}$ using continuity properties of the mapping from ξ_n to $X_{[n:n]}$. We show that there is a tractable continuous approximation for a range of distributional assumptions.

After our theoretical result, we illustrate our theory with an empirical example taken from Burke and Brown [3] who analyze commuter walking distances. We show that our theory predicts that walking distances are gamma-distributed and verify that this prediction is confirmed by the data. In the discussion section, we also propose an extension which would allow us to analyze the asymptotic behavior under an even wider range of distributional assumptions.

In Sect. 18.2 we outline the model environment. In Sect. 18.3, we provide the necessary theoretical background on point processes. Section 18.4 derives the limiting behavior of our point processes and use this to derive the limiting behavior of choice probabilities. Section 18.5 outlines the empirical application and other applications, whereas Sect. 18.6 proposes an extension to encompass a wider range of distributional assumptions. Section 18.7 concludes.

The paper is similar in aim to Malmberg and Hössjer [14]. However, they used asymptotic properties of deterministic point processes in order to analyze random utilities by methods developed in the literature on random sup measures (see O'Brien et al. [16], Resnick and Roy [18], and Stoev and Taqqu [19]). The novel approach in this paper is to analyze the problem using random point process theory instead, and this method allows for a mathematically simpler formulation than the one used in Malmberg and Hössjer [14]. Since we analyze the values X associated with the maximum U, the paper also connects to the theory of concomitants of extremes (see Ledford and Tawn [11]). The theory proposed in the extension section also relates to conditional extreme value theory, which is discussed in Heffernan and Tawn [8].

In this paper, we illustrate our theory using commuting patterns. Earlier work on random choice models with an infinite number of options has also been used to model distance dependence in international trade (Kapiarz et al. [10]). Even though the motivation for our setup comes from random choice theory, the theory has also been used in machine learning by Maddison et al. [13], who use methods in Malmberg and Hössjer [14] to derive a new way of sampling from a posterior distribution in problems of Bayesian statistics.

18.2 Model Environment

18.2.1 Model Setup and Assumptions

Consider a sequence of independent and identically distributed pairs of random variables $\{(X_i, U_i)\}_{i=1}^{\infty}$, where $X_i \in \Omega$ and $U_i \in \mathbb{R}$, and where Ω is a complete, separable, metric space.

Here X_i gives the characteristics of the choice i and Ω is the characteristic space. In case of residential choice, we might have $\Omega \subseteq \mathbb{R}^2$, where X_i gives the location of choice i. In industrial organization, $\Omega \subset \mathbb{R}^n$ might denote a multidimensional product characteristics vector, and X_i is the characteristics of a particular good. It aids intuition to think of Ω as a subset of Euclidean space, but the analysis is done

for the general case of a complete, separable, metric space. This means that the setup can be used to analyze cases where choice options are functions, for example choices of continuous consumption paths over finite intervals when the valuation is random from the econometrician's perspective.

We define $U_{n:i}$ as the ith order statistic increasing order of $\{U_1, \ldots, U_n\}$. For each n, we define the characteristic $X_{[n:i]}$ to be the X-value (the concomitant) associated with $U_{n:i}$ for a sample of size n.

We are interested in the limiting probability distribution of the characteristics of the optimal choice $X_{[n:n]}$, and to this end, we study the asymptotic behavior of the sequence of probability measures

$$C_n(\cdot) = P\left(X_{[n:n]} \in \cdot\right). \tag{18.1}$$

The distribution of (X, U) is

$$P((X, U) \in A \times B) = \int_A \mu(x; B) d\Lambda(x),$$

where $F_X = \Lambda$ is the marginal distribution of X over Ω, and $\mu(x; \cdot)$ is the regular conditional probability measure of U_i given $X_i = x$.[1] The interpretation here is that the characteristics of offers are distributed according to Λ. For example, Λ gives the distribution of potential dwellings over space in the case of residential choice, or the distribution of products over the characteristic space in case of industrial organization applications. For each offer, there is a distribution of utility $\mu(x; \cdot)$ depending on the characteristics x. We make the following assumption on μ:

Assumption 1 For the collection $\mu = \{\mu(x; \cdot); x \in \Omega\}$, there exists a function

$$p : \Omega \to (0, \infty), \tag{18.2}$$

and sequences $a_n > 0$, b_n, independent of x, and a distribution function G_α with $\alpha \in \mathbb{R}$, such that

$$\mu(x; (-\infty, a_n u + b_n])^n \to G_\alpha(u)^{p(x)} \tag{18.3}$$

as $n \to \infty$, where G_α is a distribution function of one of the following three forms:

$$G_\alpha(u) = \begin{cases} \exp(-(-u)^{-\alpha} I(u < 0)), & \alpha < 0, \\ \exp(-\exp(-u)), & \alpha = 0, \\ I(u > 0) \exp(-u^{-\alpha}), & \alpha > 0, \end{cases}$$

and $I(\cdot)$ is the indicator function.

The assumption above essentially asserts that all $\mu(x; \cdot)$ belong to the domain of attraction of the same extreme value distribution, indexed by α, and that their limiting

[1] In terms of the example in the introduction, Λ corresponds to the law of the random variables X_i, and $\mu(x; \cdot)$ corresponds to the law of the random variable $f(X_i) + \varepsilon_i | X_i = x$.

relative sizes can be described by the one dimensional parameter $p(x)$. The function $p(x)$ captures the deterministic "quality" inherent in characteristics x, which determines the limiting behavior of offer quality. The following example makes it clear in what sense $p(x)$ captures a deterministic component of utility.

Example 18.1 Assume there is a function $h(x)$ such that $\mu(x;\cdot)$ is given by an exponential distribution shifted $h(x)$ to the right. Formally, let $\mu(x;\cdot)$ be the law of a random variable $h(x)+\varepsilon$ where $\varepsilon \sim Exp(1)$.

This collection of distributions satisfies Assumption 1 when $p(x) = e^{h(x)}, a_n = 1$, $b_n = \log(n)$, and $\alpha = 0$. Equation (18.3) follows from

$$\begin{aligned} \mu(x;(-\infty, b_n + a_n u])^n &= (1 - \exp(-u - h(x) - \log n))^n \\ &\to \exp(-\exp(-u)p(x)), \\ &= G_0(u)^{p(x)}. \end{aligned} \tag{18.4}$$

The distributional assumption is not vacuous. Below is a class of distributions which does not satisfy Assumption 1.

Example 18.2 Suppose there exists a non-constant function $h(x)$ such that $\mu(x;\cdot)$ is the law of a normal distribution with mean $h(x)$ and variance 1.

Let $F \sim N(0,1)$ and assume without loss of generality that there exists an $x_0 \in \Omega$ such that $h(x_0) = 0$, and find a_n, b_n such that $F^n(a_n y + b_n)$ converges to a non-degenerate distribution function $G(y)$. Extreme value theory means that the normalization constants for a normal distribution satisfies $a_n \to 0$ and $b_n \to \infty$, and we know that the limiting distribution function $G(\cdot)$ is of Gumbel type $\alpha = 0$ (Resnick [17]).

But this means that $F^n(a_n y + b_n - h(x))$ converges to 0 if $h(x) > 0$, as

$$\begin{aligned} \lim_{n\to\infty} F^n(a_n y + b_n - h(x)) &= \lim_{n\to\infty} F^n\left[a_n\left(y - \frac{h(x)}{a_n}\right) + b_n)\right] \\ &\le \lim_{n\to\infty} F^n\left[a_n\left(y - \frac{h(x)}{a_N}\right) + b_n)\right] \\ &= G\left(y - \frac{h(x)}{a_N}\right) \end{aligned}$$

for any sufficiently large N. Let $N \to \infty$ and we obtain the conclusion. As the limit is 0, we need $p(x) = \infty$ which violates that $p(x) < \infty$. On the other hand, we can use an analogous reasoning to conclude that $F^n(a_n y + b_n - h(x))$ converges to 1 if $h(x) < 0$, so that $p(x) = 0$. This violates $p(x) > 0$.

We conclude that a non-constant function $h(x)$ is not consistent with Assumption 1.

It limits the theory that the traditional normal regression structure does not satisfy Assumption 1. The reason is that the normal distribution is too thin-tailed. Formally, the condition for when the linear regression formulation works is whether the limit

$$\lim_{u\to\infty} \frac{P(U + h(x_1) > u)}{P(U + h(x_2) > u)}$$

exists and is not 0 or ∞ when $h(x_1) \neq h(x_2)$. This condition holds when U is exponentially distributed but not when U is normally distributed. When U is normally distributed, the limit is ∞ for $h(x_1) > h(x_2)$ and 0 for $h(x_1) < h(x_2)$. In Sect. 18.6, we propose an extension which would allow us to analyze normal regression functions.

18.2.2 Point Process Formulation and Strategy

The sequence $\{(X_i, U_i)\}_{i=1}^n$ can be viewed as a random collection of points in $\Omega \times \mathbb{R}$, and can be described as a sequence of point processes ξ_n. We will show that after a suitable transformation, this sequence of point processes ξ_n converges to a Poisson point process ξ in a sense which will be formalized later. As

$$C_n(A) = P(X_{[n:n]} \in A) = P\left(\sup_{i:X_i\in A} U_i > \sup_{i:X_i\notin A} U_i\right)$$

is a functional on our point process ξ_n, the problem of finding $\lim_{n\to\infty} C_n$ reduces to determine whether this functional is continuous. In this case, we can use the limiting point process ξ to calculate our results.

We will start with an introduction to point processes – in particular sufficient conditions for convergence. After this, we will apply the point process machinery to our setup, and characterize the limit of our point process. Once this is done, we will define random fields taking point processes as inputs, and derive the asymptotic behavior of C_n from continuity properties of these random fields.

18.3 Background on Point Processes and Convergence Results

This section contains background results and a notational machinery for point processes. See Chapter 3 of Resnick [17] for a more detailed treatment.

Throughout this discussion, the generic point process will take values in a locally compact set E, with an associated σ-algebra $\mathcal{E}$. For the purpose of our discussion, E will be a subset of $\Omega \times \mathbb{R}$, and we assume that $\mathcal{E} = \mathcal{B}(E)$ is the Borel σ-algebra. A *point mass* is a set function, defined by

$$\delta_z(F) = \begin{cases} 1, & \text{if } z \in F, \\ 0, & \text{if } z \notin F, \end{cases}$$

where $F \subseteq E$, $F \in \mathcal{E}$. A *point measure* is a measure $m(\cdot)$ on E such that there exists a countable collections of points $\{z_k\} \subseteq E$ and numbers $\{w_k\} \geq 0$, such that

$$m(\cdot) = \sum_{z_k} w_k \delta_{z_k}(\cdot).$$

We will confine our attention to the case $w_k \equiv 1$. Let $\mathcal{M}_P(E)$ be the set of point measures on E, and let it have the minimal σ-algebra which makes

$$\{m \in \mathcal{M}_P(E) : m(F) \in B\}$$

measurable for all $F \in \mathcal{E}$, $B \in \mathcal{B}(\mathbb{R})$ where $m(F)$ is the point measure m evaluated at the set F and $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$. We define a *point process* to be a random element of $\mathcal{M}_P(E)$.

If N is an arbitrary point process, we define the Laplace transform ψ associated with N as

$$\begin{aligned} \psi_N(f) &= E\left(\exp\left\{-\int_E f(x)N(dx)\right\}\right) \\ &= \int_{M_P(E)} \exp\left\{-\int_E f(x)m(dx)\right\} P^N(dm). \end{aligned} \tag{18.5}$$

Here P^N is a probability measure over the set $\mathcal{M}_P(E)$ which corresponds to the distribution of N. Moreover, the class of functions f for which we are interested in ψ_N is usually the continuous non-negative functions on E with a compact support. We write $C_K^+(E)$ to denote this set.

Definition 18.1 A sequence of point processes N_n, $n \geq 0$, converges in a point process sense to N_0, written $N_n \Rightarrow_p N_0$, if

$$\psi_{N_n}(f) \to \psi_{N_0}(f)$$

for all $f \in C_K^+(E)$.

We use the notation $\Longrightarrow$ for weak convergence of vector valued random variables in Euclidean space or on Ω, in contrast to $\Rightarrow_p$ for point process convergence.

Definition 18.2 Let E be a metric space. We call $F \subseteq E$ relatively compact if its closure $\bar{F}$ in E is compact.

Definition 18.3 Let μ be a measure on a metric space $\mathcal{X}$. We say that a sequence of measures μ_n converges *vaguely* to μ, written

$$\mu_n \Rightarrow_v \mu,$$

if

$$\mu_n(F) \to \mu(F)$$

for all relatively compact F with $\mu(\partial F) = 0$, where ∂F is the boundary of the set F.

Definition 18.4 For a point process N, the Laplace functional associated with N is defined by

$$\Psi_N(f) = E\left[\exp(-N(f))\right]$$

where $N(f) = \sum_{x \in N} f(x)$.

It is known from point process theory that the Laplace functional uniquely defines a point process. Thus, the Laplace functional can be used to define a Poisson process and derive its properties (see, for example, Resnick [17], p. 130).

Definition 18.5 A Poisson process with intensity measure μ is a point process defined by the Laplace functional

$$\Psi_N(f) = e^{-\int_E (1-e^{-f(x)}) d\mu(x)}.$$

Proposition 18.1 *For any $F \in \mathcal{E}$, and any non-negative integer k, a Poisson process satisfies*

$$P(N(F) = k) = \begin{cases} e^{-\mu(F)}(\mu(F))^k/k!, & \text{if } \mu(F) < \infty, \\ 0, & \text{if } \mu(F) = \infty, \end{cases}$$

and that for any $k \geq 1$, if $F_1, \ldots, F_k$ are mutually disjoint sets in $\mathcal{E}$, then $\{N(F_i)\}$ are independent random variables.

Our main theorem will also depend on the following proposition which is a modification of a result presented in the proof of a more extensive Proposition 3.21 in Resnick [17].

Proposition 18.2 *For each n, suppose $\{Z_{n,j} : 1 \leq j \leq n\}$ are independent and identically distributed (i.i.d.) random variables on E and that*

$$nP(Z_{n,1} \in \cdot) \Rightarrow_v \mu.$$

where μ is a measure on E. Then

$$N_n = \sum_{j=1}^{n} \delta_{Z_{n,j}} \Rightarrow_p N$$

where N is a Poisson random measure on E with intensity μ.

Proof This proof is essentially equivalent to the first half of the proof of Proposition 3.21 in Resnick [17]. We use that convergence in point measures is equivalent to convergence in Laplace functionals. Indeed, pick an arbitrary $f \in C_K^+(E)$, with a compact support $F \subseteq E$. Then:

$$\begin{aligned}
\psi_{N_n}(f) &= E\exp\{-N_n(f)\} \\
&= E\exp\left\{-\textstyle\sum_{j=1}^{n} f(Z_{n,j})\right\} \\
&= \left(E\exp\{-f(Z_{n,1})\}\right)^n \\
&= \left(1-\frac{\int_E (1-e^{-f(z)})nP[Z_{n,1}\in dz]}{n}\right)^n \\
&= \left(1-\frac{\int_F (1-e^{-f(z)})nP[Z_{n,1}\in dz]}{n}\right)^n \\
&\to e^{-\int_F (1-e^{-f(z)})d\mu(z)} \\
&= e^{-\int_E (1-e^{-f(z)})d\mu(z)} \\
&= \psi_N(f),
\end{aligned} \tag{18.6}$$

where the convergence step is obtained from the vague convergence of $nP[Z_{n,1} \in \cdot]$. Indeed, vague convergence is equivalent to weak convergence on every compact subspace. As $1 - e^{-f(z)}$ continuous and bounded, and weak convergence means that the integral of every continuous and bounded function converges, we get the desired result. Thus,

$$N_n \Rightarrow_p N$$

as required. □

Before giving the full proof of Theorem 18.1, we state and prove the following lemma:

Lemma 18.1 *If $(\mathcal{X} \times \mathcal{U}, \Lambda \times \nu)$ is a product measure space, where $\mathcal{X}$ and $\mathcal{U}$ are two complete, separable metric spaces, and if $F \subseteq \mathcal{X} \times \mathcal{U}$ satisfies*

$$(\Lambda \times \nu)(\partial F) = 0,$$

then

$$\nu(\partial F_x) = 0 \quad \Lambda - a.e.$$

where $F_x = \{u \in \mathcal{U} : (x, u) \in F\}$ is the cross-section of F at the point x, and a.e. refers to convergence almost everywhere (or almost surely).

Proof We note that if we write

$$B = \{(x, u) \in \mathcal{X} \times \mathcal{U} : u \in \partial F_x\},$$

we have

$$B \subseteq \partial F$$

(as each ball around a point $(x, u) \in B$ contains both a point within and outside F). Thus, as

$$(\Lambda \times \nu)(B) = \int_{\mathcal{X}} \nu(\partial F_x) d\Lambda(x) \leq (\Lambda \times \nu)(\partial F) = 0,$$

we get that $\nu(\partial F_x) = 0$ Λ-almost everywhere. □

18.4 Limiting Behavior of Choice Probabilities

In this section, we use point process theory to derive the limit of the choice probabilities $C_n(\cdot) = \mathbb{P}(X_{[n:n]} \in \cdot)$. We first show that the point process generated by the collection $\{(X_i, U_i)\}$ converges to a Poisson process after a suitable transformation. We then use this fact to calculate the limit of C_n.

18.4.1 Convergence of Point Process

We consider a sequence of transformations

$$g_n(u) = (u - b_n)/a_n$$

where a_n, b_n are chosen to ensure extreme value convergence for all $x \in \Omega$ as in (18.3) of Assumption 1.

Let $\delta_{(x,u)}$ denote a one point distribution at (x, u) and define the extremal marked point process (cf. Resnick [17])

$$\xi_n = \sum_{i=1}^{n} \delta_{(X_i, g_n(U_i))} \tag{18.7}$$

for a sample of size n. This is a point process on $(\Omega \times \mathbb{R}, \mathcal{B}(\Omega \times \mathbb{R}))$.

We are now ready to formulate our first main result. It states that ξ_n converges to a Poisson process with a product intensity measure which multiplies the initial measure Λ on Ω with $p(x)$. Before stating this result, we first introduce a few concepts.

Definition 18.6 A random variable X stochastically dominates a random variable Y if

$$\mathbb{P}(X \geq x) \geq P(Y \geq x) \quad \forall x \in \mathbb{R}.$$

We also say that a measure μ_X on the real numbers dominates μ_Y if they are laws of random variables X and Y and X stochastically dominates Y.

Theorem 18.1 *Let G_α and p be as in Assumption 1. Suppose that for each compact subset $A \subseteq \Omega$, the function $p : \Omega \to (0, \infty)$ is bounded on that subset, and that $\mu(x_0, \cdot)$, for some $x_0 \in \Omega$ is an upper bound for all $\{\mu(x; \cdot);\ x \in A\}$ in the sense of stochastic dominance. Then*

$$\xi_n \Rightarrow_p \xi,$$

as $n \to \infty$ where ξ_n is given by (18.7), and ξ is a Poisson random measure on $(\Omega \times \mathbb{R}, \mathcal{B}(\Omega \times \mathbb{R}))$ with mean intensity $\Lambda_p \times \nu_\alpha$, where

$$\Lambda_p(A) = \int_A p(x)\Lambda(dx)$$

for all relatively compact $A \in \mathcal{B}(\Omega)$ *and*

$$v_\alpha([u,\infty)) = -\log(G_\alpha(u)) = \begin{cases} I(u<0)(-u)^{-\alpha}, & \text{if } \alpha < 0 \text{ and } u < 0\,, \\ \exp(-u), & \text{if } \alpha = 0, \\ u^{-\alpha}, & \text{if } \alpha > 0 \text{ and } u > 0. \end{cases}$$

Proof Note that we have $G_\alpha(u) = 0$ for $\alpha > 0$ and $u \leq 0$. Whenever $\alpha > 0$, it is therefore implicit in the proof that $u > 0$. Using the proof of Proposition 18.2, it suffices to show that

$$nP((X_1, g_n(U_1)) \in \cdot) \Rightarrow_v \Lambda_p \times v_\alpha,$$

i.e. that

$$nP((X_1, g_n(U_1)) \in F) \to (\Lambda_p \times v_\alpha)(F),$$

for all $F \subseteq \Omega \times \mathbb{R}$ which are relatively compact sets with respect to $\mathcal{B}(\Omega \times \mathbb{R})$ and satisfy

$$(\Lambda_p \times v_\alpha)(\partial F) = 0.$$

Henceforth, let F be an arbitrary set with these properties. Now, we note that

$$nP((X_1, g_n(U_1)) \in F) = \int_\Omega nP(g_n(U_1) \in F_x | X_1 = x) d\Lambda(x),$$

where F_x is the x-cross section of F. Thus, our task is to show that

$$\int_\Omega nP(g_n(U_1) \in F_x | X_1 = x) d\Lambda(x) \to \int_\Omega p(x) v_\alpha(F_x) d\Lambda(x).$$

We do this first by showing that the integrand converges almost everywhere to the desired quantity, and then we show that the sequence of integrands satisfy regularity conditions allowing us to infer convergence of integrals from pointwise convergence.

We observe that for every x,

$$nP(g_n(U_1) \in \cdot | X_1 = x) \Rightarrow_v p(x) v_\alpha(\cdot). \tag{18.8}$$

Indeed, it is true that for any sequence x_n such that

$$(x_n)^n \to a, \tag{18.9}$$

we have

$$n(1 - x_n) \to -\log(a). \tag{18.10}$$

Thus, letting $x_n = P(g_n(U_1) < u | X_1 = x)$ and using Assumption 1, we obtain

$$nP(g_n(U_1) \geq u | X_1 = x) \to -p(x)\log(G_\alpha(u)) = p(x)\nu_\alpha([u, \infty)). \quad (18.11)$$

In order to deduce (18.8) from (18.11), we can note that if we have a measure γ with

$$\gamma([u, \infty)) < +\infty$$

for some u, then vague convergence of γ_n to γ is equivalent to

$$\gamma_n([u, \infty)) \to \gamma([u, \infty)), \quad (18.12)$$

for all u such that $\gamma(\{u\}) = 0$. This can be seen by noting that if (18.12) is true, then the sequence $P_{nu}(\cdot) = \gamma_n(\cdot \cap [u, \infty))/\gamma_n([u, \infty))$ of probability measures converges weakly for all continuity points u of $\gamma([u, \infty))$ to $P_u(\cdot) = \gamma(\cdot \cap [u, \infty))/\gamma[u, \infty))$, and hence $P_{nu}(F) \to P_u(F)$ for all such u, from which (18.8) follows.

Now, using the previous lemma, we know that

$$\nu_\alpha(\partial F_x) = 0 \quad \Lambda_p - a.e.,$$

which means that

$$p(x)\nu_\alpha(\partial F_x) = 0 \quad \Lambda_p - a.e.$$

as $p(x) > 0$ implies that $p(x)\nu_\alpha$ and ν_α are equivalent for all $x \in \Omega$. Thus, we can use (18.8) to conclude that

$$nP(g_n(U_1) \in F_x | X_1 = x) \to p(x)\nu_\alpha(F_x) \quad \Lambda_p - a.e.$$

Therefore, we have established pointwise convergence of the integrand almost everywhere.

Now, we seek to show that $nP(g_n(U_1) \in F_x | X_1 = x)$ is uniformly bounded over n and Ω to ensure that pointwise convergence almost everywhere implies convergence in integrals. To do so, we want to define a maximal random variable which dominates $nP(g_n(U_1) \in F_x | X_1 = x)$ for all n and x.

We write

$$\pi_\Omega : (x, u) \mapsto x$$

and

$$\pi_{\mathbb{R}} : (x, u) \mapsto u$$

for the projection on Ω and $\mathbb{R}$ respectively. In this case, we know that $\pi_\Omega(F)$ and $\pi_{\mathbb{R}}(F)$ are relatively compact sets of Ω and $\mathbb{R}$ respectively. By the assumptions in the theorem, there is an $x_0(F) \in \Omega$ that maximizes p on $\pi_\Omega(F)$. This means that a random variable $\bar{U}(F)$ with measure $\mu(x_0(F); \cdot)$ dominates $U_1 | X_1 = x$ stochastically for all $x \in \pi_\Omega(F)$. Write $\bar{p}(F) = p(x_0(F))$ for the corresponding p-value

of p. Furthermore, we can define $\underline{u}$ as the smallest u-value attained on the whole set $\pi_{\mathbb{R}}(F)$, which again is finite by the assumption of F being relatively compact. Combining these two definitions gives us

$$\begin{aligned} nP(g_n(U_1) \in F_x | X_1 = x) &\leq nP(g_n(U_1) \geq \underline{u} | X_1 = x) \\ &\leq nP(g_n(\bar{U}(F)) \geq \underline{u} | X_1 = x) \\ &= nP(g_n(\bar{U}(F)) \geq \underline{u}) \\ &\to \bar{p}(F)\nu_\alpha([\underline{u}, \infty)) \\ &< +\infty, \end{aligned}$$

which means that $nP(g_n(U_1) \in F_x | X_1 = x)$ is uniformly bounded. Using the bounded convergence theorem, we get

$$\begin{aligned} nP((X_1, g_n(U_1)) \in F) &= \int_\Omega nP(g_n(U_1) \in F_x | X_1 = x) d\Lambda(x) \\ &\to \int_\Omega \nu_\alpha(F_x) p(x) d\Lambda(x) \\ &= (\Lambda_p \times \nu_\alpha)(F), \end{aligned}$$

which completes the proof. □

This theorem is similar to Proposition 3.21 in Resnick [17]. There are two differences. First, in [17], the author considers $\xi_n = \sum_{j=1} \delta_{(jn^{-1}, g_n(U_j))}$ where $\{U_j\}$ is a sequence of independent and identically distributed random variables. Thus, the difference is that we model the first coordinate as a random variable, and let the distribution of the second coordinate depend on this first coordinate. Furthermore, we let X take values in a general separable metric space. The differences add some technicalities to the proof, but they turn out not to affect the main result.

We can also note that the distributional assumptions ensure that the optimal choice and the maximum value are independent in the limit, which means that we can write the product measure as a direct product of measures on the two spaces. See Fosgerau et al. [6] for a general discussion of probability distributions having this invariance property.

The assumption that p is bounded on compact sets is for example satisfied whenever p is continuous. The assumption that we can construct a stochastically dominating random variable for each compact set is a technical assumption required to apply the bounded convergence theorem. As a counterexample when the theorem fails, consider the model of Example 18.1, with Λ having a uniform distribution on $\Omega = [0, 1)$, $h(x) = -\log(x)$ and $p(x) = x^{-1}$ for $x \neq 0$, whereas $h(0) = 0$ and $p(0) = 1$. In order to have convergence $\xi_n(F) \Longrightarrow \xi(F)$ for relatively compact sets $F \in \mathcal{E}$ with $\mu(\partial F) = 0$, for which the closure of the projection of F onto Ω does not contain 0, we take $a_n = 1$ and $b_n = \log(n)$. On the other hand, if $F = [0, \delta] \times [-K, K]$, it can be seen that $\xi_n(F)$ tends to infinity with probability 1 as $n \to \infty$, for any values of $0 < \delta < 1$ and $K > 0$.

18.4.2 Convergence of Choice Probability Distribution

Recall that our task is to study the limiting behavior of C_n as defined in (18.1). The key to connect this limit to point processes is the observation that because g_n is strictly increasing for all n: $C_n(A) = P(X_{[n:n]} \in A) = P(M_{\xi_n}(A) > M_{\xi_n}(A^c))$ for all $A \in \mathcal{B}(\Omega)$, with M_{ξ_n} a random field defined as $M_{\xi_n}(A) = \max_{\substack{X_i \in A \\ 1 \le i \le n}} g_n(U_i)$, $A \in \mathcal{B}(\Omega)$, where $\mathcal{B}(\Omega)$ is the Borel sigma algebra over Ω, and ξ_n is the point process from (18.7). This formulation of the argmax-measure C_n in terms of random fields defined over point processes allows us to generalize the notion of argmax to the limiting case where the number of offers goes to infinity. We will study the limiting behavior of finite-dimensional distributions of M_{ξ_n}. This will allow us to calculate the limit of C_n.

The mean intensity $\Lambda_p \times \nu_\alpha$ in Theorem 18.1 is a non-finite measure defined over $\Omega \times \mathbb{R}$. However, if $\Lambda(\Omega) < \infty$, it is possible to write $\Omega \times \mathbb{R}$ as a countable union of sets with finite measures $\Lambda_p \times \nu_\alpha$. Hence, the realization of the point process ξ has countably infinite many points almost surely. If we write $\{X_i^\infty, U_i^\infty\}_{i=1}^\infty$ for the sequence of random variables giving the locations of these points, we can define, $M_\xi(A) = \max_{i; X_i^\infty \in A} U_i^\infty$ as a random field giving the highest variable attained for a given set $A \subseteq \Omega$, and $C(A) = P(M_\xi(A) > M_\xi(A^c))$ for the probability that A will contain the largest U-element.

Proposition 18.3 *If $\Lambda_p(\Omega) < \infty$, we have $C(A) = \Lambda_p(A)/\Lambda_p(\Omega)$.*

Proof Suppose first that $\Lambda_p(A^c) = 0$ or $\Lambda_p(A) = 0$. In this case, it is clear that we have $C(A) = 1$ or $C(A) = 0$ respectively as required by the formula for $A \in \mathcal{B}(\Omega)$. Indeed, using the convention that the supremum of an empty set is minus infinity, if $\Lambda_p(A) = 0$, then $M_\xi(A) = -\infty$ almost surely. As $M_\xi(A^c) > -\infty$ almost surely, we will get $C(A) = 0$. A similar reasoning applies to A^c.

Furthermore, since ξ is a Poisson random measure with mean measure $\Lambda_p \times \nu_\alpha$, we note that if $\Lambda_p(\Omega) < \infty$ we have that $M_\xi(A)$ and $M_\xi(A^c)$ are two independent, proper random variables with

$$P(M_\xi(A) \le y) = P(\xi(A \times (y, \infty)) = 0) = e^{-\Lambda_p(A)\nu_\alpha((y,\infty))} \tag{18.13}$$

$$P(M_\xi(A^c) \le y) = P(\xi(A^c \times (y, \infty)) = 0) = e^{-\Lambda_p(A^c)\nu_\alpha((y,\infty))}. \tag{18.14}$$

Standard calculations yield

$$P(M_\xi(A) > M_\xi(A^c)) = \frac{\Lambda_p(A)}{\Lambda_p(A) + \Lambda_p(A^c)} = \Lambda_p(A)/\Lambda_p(\Omega)$$

and the proof is complete. □

From this result, we automatically get that C is a probability measure as it is a normalized version of Λ_p which is a finite measure.

In order to prove that C_n converges weakly, we need some additional results. We use that

$$\nu_1 \ll \mu_1 \text{ and } \nu_2 \ll \mu_2 \Rightarrow \nu_1 \times \nu_2 \ll \mu_1 \times \mu_2, \tag{18.15}$$

where $\ll$ means "absolutely continuous with respect to".

We will also use that if ξ_n are point processes, ξ is a Poisson process, and

$$\xi_n \Rightarrow_p \xi,$$

then

$$P(\xi_n(F) = 0) \to P(\xi(F) = 0) \tag{18.16}$$

for all $F \in \mathcal{E}$ with $\mu(\partial F) = 0$, where μ is the intensity measure of ξ.

After these preliminaries, we are ready to state our second main result:

Theorem 18.2 *If $\Lambda_p(\Omega) < \infty$, we have*

$$C_n(\cdot) \Rightarrow C(\cdot) = \frac{\Lambda_p(\cdot)}{\Lambda_p(\Omega)}. \tag{18.17}$$

Proof Assume we have A with $C(\partial A) = 0$. We aim to prove that $C_n(A) \to C(A)$. By Proposition 18.3, C and Λ_p are equivalent, and we have $\Lambda_p(\partial A) = 0$. Noting that the result is clearly true whenever $\Lambda_p(A) = 0$ or $\Lambda_p(A^c) = 0$, we can assume that both are different from 0. By (18.13) and (18.14), this means that $(M_\xi(A), M_\xi(A^c))$ is a proper random variable on $\mathbb{R}^2$, and we will show that $(M_{\xi_n}(A), M_{\xi_n}(A^c))$ jointly converge weakly to this random variable. Indeed, consider

$$\begin{aligned} P(M_{\xi_n}(A) \le x_1, M_{\xi_n}(A^c) \le x_2) &= P\left(\xi_n(A \times (x_1, \infty) \cup A^c \times (x_2, \infty)) = 0\right) \\ &\to P\left(\xi(A \times (x_1, \infty) \cup A^c \times (x_2, \infty)) = 0\right) \\ &= P(M_\xi(A) \le x_1, M_\xi(A^c) \le x_2) \\ &= F_{M_\xi(A), M_\xi(A^c)}(x_1, x_2). \end{aligned}$$

The convergence step uses (18.16) and that

$$\partial\left(A \times (x_1, \infty) \cup A^c \times (x_2, \infty)\right) \subset \partial A \times (\min(x_1, x_2), \infty) \cup A \times (\{x_1\} \cup \{x_2\})$$

and we have $(\Lambda_p \times \nu_\alpha)(\partial A \times (\min(x_1, x_2), \infty) \cup A \times (\{x_1\} \cup \{x_2\})) = 0$ since $\Lambda_p(\partial A) = 0$ and $\nu_\alpha(\{x_1\} \cup \{x_2\}) = 0$, where $\Lambda_p \times \nu_\alpha$ is the intensity measure of ξ. Hence

$$(M_{\xi_n}(A), M_{\xi_n}(A^c)) \Rightarrow (M_\xi(A), M_\xi(A^c)).$$

Defining

$$D = \{(a, b) \in \mathbb{R}^2 : a > b\}$$

and using (18.15), with $\nu_1 \sim M_\xi(A)$, $\nu_2 \sim M_\xi(A^c)$, and μ_1, μ_2 Lebesgue measure in $\mathbb{R}$, to conclude that

$$P((M_\xi(A), M_\xi(A^c)) \in \partial D) = 0$$

we get

$$\begin{aligned} C_n(A) &= P(M_{\xi_n}(A) > M_{\xi_n}(A^c)) \\ &= P((M_{\xi_n}(A), M_{\xi_n}(A^c)) \in D) \\ &\to P((M_\xi(A), M_\xi(A^c)) \in D) \\ &= C(A) \end{aligned}$$

and the proof is complete. □

18.5 Examples

Here we provide a few examples to illustrate our theory.

Example 18.3 (*Exponential and mixture models*) This example extends Example 18.1, and calculates the argmax distribution associated with that example. Consider a family of models where the regular conditional probability measure $\mu(x; \cdot)$ is indexed by α, and where for each $A \in \mathcal{B}(\mathbb{R})$ we have

$$\mu_\alpha(x; A) = \begin{cases} P\left((2 \times 1_{\{V_1 < p(x)\}} - 1)(1 - V_2^{-1/\alpha}) \in A\right), & \alpha < 0, \\ P\left(\log(p(x)/V_1) \in A\right), & \alpha = 0, \\ P\left((2 \times 1_{\{V_1 < p(x)\}} - 1)V_2^{-1/\alpha} \in A\right), & \alpha > 0, \end{cases}$$

where $V_1, V_2 \sim U(0, 1)$ are two independent and uniformly distributed random variables on $(0, 1)$. A bit less formal, we may write

$$\mu_\alpha(x) \sim \begin{cases} -(1 - p(x))\text{Beta}(1, -\alpha) + p(x)\text{Beta}(1, -\alpha), & \alpha < 0, \\ \text{Exp}(\log(p(x)), 1), & \alpha = 0, \\ -(1 - p(x))\text{Pareto}(\alpha, 1) + p(x)\text{Pareto}(\alpha, 1), & \alpha > 0, \end{cases}$$

where $\text{Beta}(a, b)$ refers to a beta distribution with density $Cx^{a-1}(1 - x)^{b-1}$ on $(0, 1)$, $\text{Exp}(a, b)$ is a shifted exponential distribution with location parameter a and scale parameter b, having distribution function $1 - e^{-(x-a)/b}$ for $x \geq a$, $\text{Pareto}(\alpha, b)$ is a Pareto distribution with shape parameter α and scale parameter b, corresponding to a distribution function $1 - (x/b)^{-\alpha}$ for $x \geq b$. We let $x_0 \in \Omega$ be an arbitrary point for which $p(x_0) = 1$.

We have chosen the parameter α for $\mu_\alpha(x, \cdot)$ in a way so that (18.3) holds, with $a_n = n^{1/\alpha}$, $b_n = 1$ when $\alpha < 0$, $a_n = 1$, $b_n = \log(n)$ when $\alpha = 0$, and $a_n = n^{1/\alpha}$, $b_n = 0$ when $\alpha > 0$. When $\alpha = 0$, this follows from tail properties of the exponential

distribution, as shown in Example 18.1. For $\alpha \neq 0$, we have that

$$\begin{aligned}\mu_\alpha(x; (-\infty, b_n + a_n u])^n &= \{1 - p(x)(1 - \mu_\alpha(x_0; (-\infty, b_n + a_n u]))\}^n \\ &\to G_\alpha(u)^{p(x)}.\end{aligned}$$

In the last step we used that $\mu_\alpha(x_0, (-\infty, b_n + a_n u])^n \to G_\alpha(u)$. This is a well known fact of univariate extreme value theory (see for instance Fisher and Tippett [5], Gnedenko [7], and Chapter 1 in Resnick [17]), and it follows from tail properties of the beta and Pareto distributions.

This means that for all these three families of distribution, the choice probabilities will give us a tilted distribution $p \times \Lambda$ which modifies the underlying Λ-distribution with p. This effect captures that areas with a high deterministic utility component p are relatively more likely to get chosen. For the case $\alpha = 0$ this effect means that if utility is given by $U_i = h(x_i) + \varepsilon_i$, where $\varepsilon_i \sim Exp(1)$, then the choice distribution is an exponential tilt $e^{h(x)}\Lambda(dx)$ of the original distribution.

Example 18.4 (*An example from the commuting literature*) If we focus on $\alpha = 0$ in the previous example, we have an interesting special case. Suppose that a person has received a new job, and potential residencies are distributed uniformly on $B(0, R)$, a disk in $\mathbb{R}^2$. There is a linear cost $c||x||$ associated with travelling to a location $x \in B(0, R)$, and there is an exponentially distributed random component associated with each residence. This means that utility is given by $U|X = x \sim \text{Exp}(-c||x||, 1)$, where $||x||$ is the Euclidean distance from the origin. This gives a very simple model to think about commuting choices. In this case, Λ has a uniform distribution on $B(0, R)$, and $p(x) = \text{Exp}(-c||x||)$. Thus, we get $C(A) = \dfrac{\int_A e^{-c||x||}dx}{\int_{B(0,R)} e^{-c||x||}dx}$. The particular direction of commuting is often not as interesting as the distribution of distances. The probability that we commute less than r is given by

$$C(\{x : ||x|| \leq r\}) = \frac{\int_0^r se^{-cs}ds}{\int_0^R se^{-cs}ds},$$

which we recognize as a truncated Gamma$(2, 1/c)$-distribution.

There is suggestive evidence that travel patterns follow a gamma distribution over short distances. One good source is Burke and Brown [3], which documents the distances people walk for transport purposes to different destinations. The data was collected from a survey in Brisbane. Even though the investigators not only measured the time walked to work, the situation is somewhat analogous to the example above in that walking is a roughly linear cost. They found that the distance walked for one-leg trips is very close to a gamma distribution with shape parameter α and scale parameter β, and the same for the total distance walked from train stations to end destinations (see Figs. 18.1 and 18.2).

We see that the estimated parameters $(\hat{\alpha}, \hat{\beta})$ are $(1.42, 0.66)$ and $(2.13, 0.37)$ respectively. The estimated shape parameter is close to but not exactly 2 as would be

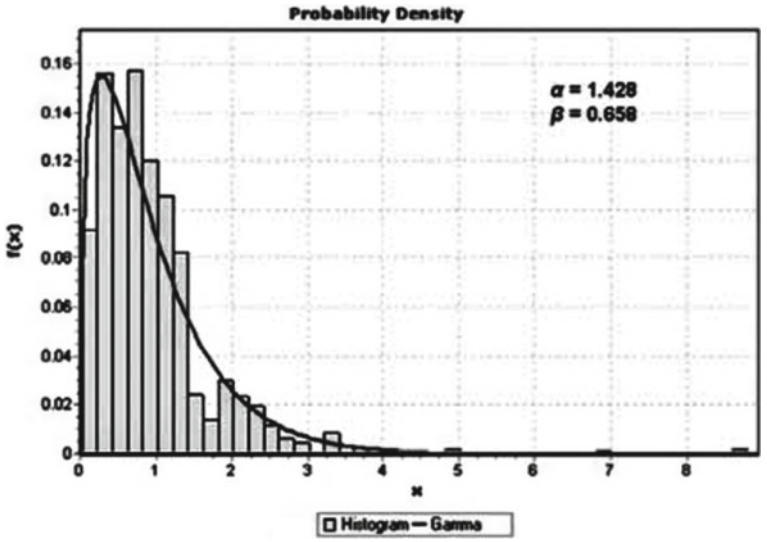

Fig. 18.1 Histogram over walking distances to final destination and fitted Gamma(α, β)-distribution (Burke and Brown [3])

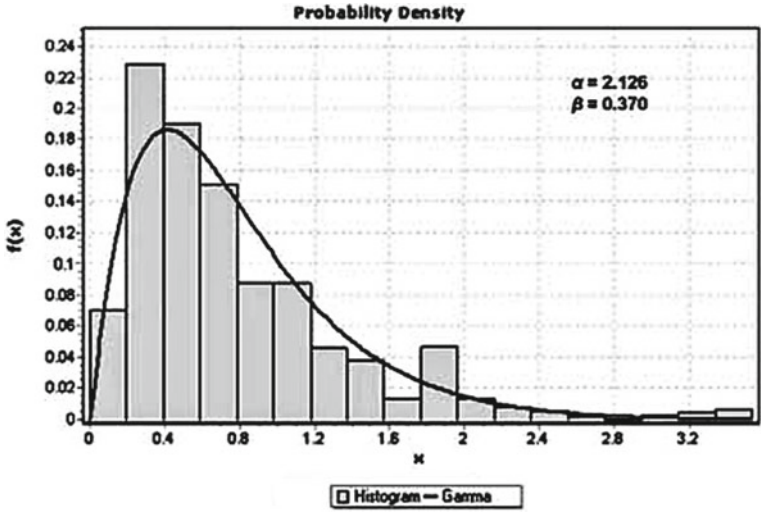

Fig. 18.2 Histogram over walking distances from train station to final destination and fitted Gamma(α, β)-distribution (Burke and Brown, [3])

predicted by the theory. The focus in the paper is to test the distributional assumption rather than to find the exact parameters, and the authors report an Anderson-Darling test but no standard errors on the parameter estimates. Hence, we do not know if α is significantly different from 2.

Example 18.5 (*The logit model: a special case*) Let Λ be a uniform distribution on the finite support $\{x_1, \ldots, x_{n_0}\}$. As in Example 18.1, let utilities be given by

$$U_j | X_j = x \sim \text{Exp}(h(x), 1). \tag{18.18}$$

This corresponds to $p(x_i) = e^{h(x_i)}$ and we get

$$C(\{x_i\}) = \frac{e^{h(x_i)}}{\sum_{j=1}^{n_0} e^{h(x_j)}}.$$

This corresponds to the famous logit model from the random choice literature (McFadden [15]).

The following is an example where we let Ω be a functional space. This shows that the methodology can be applied to more general spaces than subsets of Euclidean space, and motivates the more general space definition we introduced in Sect. 18.2.

Example 18.6 Let Ω be the space of bounded functions on $[0, 1]$, metrized by the sup-norm. A function $x \in \Omega$ describes a continuum of choice characteristics. An agent values a function $x \in \Omega$ by sampling k points of $[0, 1]$ according to a density function g, and valuing them according to their sum and an exponentially distributed noise term on each observation. In this case X is a random variable taking values in Ω with law Λ. Algebraically,

$$\begin{aligned} U | X = x &= \frac{\sum_{j=1}^{k} x(T_j)}{k} + \frac{\sum_{j=1}^{k} \varepsilon_j}{k} \\ &= h_k(x) + \varepsilon \end{aligned}$$

where T_j are i.i.d. distributed on $[0, 1]$ with density function g, and $\varepsilon_j \sim \text{Exp}(1)$ independently. We want to find the argmax distribution on Ω. We will treat a sequence of approximations as equalities, and verify ex post that such a treatment is justified.

The random variable ε has a Gamma$(k, 1/k)$ distribution, which means that

$$\bar{F}_\varepsilon(z) \equiv 1 - F_\varepsilon(z) = \sum_{m=0}^{k-1} e^{-kz} \frac{(kz)^m}{m!} \sim \frac{e^{-kz}(kz)^{k-1}}{(k-1)!},$$

where the ratio of the last two expressions tends to 1 when z gets large. Now, we use that x is bounded to get $\underline{y}(x) \leq \inf_{t \in [0,1]} x(t) \leq \sup_{t \in [0,1]} x(t) \leq \bar{y}(x)$. We write $\mu(x; \cdot)$ for the law of $U | X = x$, and approximate the upper tail when u is large:

$$
\begin{aligned}
1-\mu\left(x;(-\infty,u]\right) &= 1-\int_{\underline{y}(x)}^{\bar{y}(x)} F_\varepsilon(u-y)dF_{h_k(x)}(y) \\
&= \int_{\underline{y}(x)}^{\bar{y}(x)} \bar{F}_\varepsilon(u-y)dF_{h_k(x)}(y) \sim \int_{\underline{y}(x)}^{\bar{y}(x)} \frac{(k(u-y))^{k-1}}{(k-1)!}e^{-k(u-y)}dF_{h_k(x)}(y) \\
&\sim \frac{(ku)^{k-1}}{(k-1)!}\int_{\underline{y}(x)}^{\bar{y}(x)} e^{-k(u-y)}dF_{h_k(x)}(y) \sim \bar{F}_\varepsilon(u)p_k(x),
\end{aligned}
$$

where $p_k(x)$ is the moment generating function of $F_{h_k(x)}$ with argument k. We write $\eta_k(u) = \bar{F}_\varepsilon(u)p_k(x) - (1-\mu\left(x;(-\infty,u]\right))$ for the approximation error. Now define $a_n = 1/k$ and $b_n = \bar{F}_\varepsilon^{-1}(1/n)$, which gives us

$$
\begin{aligned}
\mu(x;(-\infty,a_nu+b_n])^n &= \left(1-\bar{F}_\varepsilon\left(\tfrac{u}{k}+b_n\right)p_k(x)+\eta_k\left(\tfrac{u}{k}+b_n\right)\right)^n \\
&\sim \left(1-\bar{F}_\varepsilon\left(\tfrac{u}{k}+b_n\right)p_k(x)\right)^n \\
&\sim \left(1-\frac{e^{-k\left(\frac{u}{k}+b_n\right)}k^{k-1}\left(\frac{u}{k}+b_n\right)^{k-1}}{(k-1)!}p_k(x)\right)^n \\
&\sim \left(1-p_k(x)e^{-u}\bar{F}_\varepsilon(b_n)\right)^n \\
&= \left(1-\tfrac{p_k(x)e^{-u}}{n}\right)^n \\
&\to (e^{-e^{-u}})^{p_k(x)} = G_0(u)^{p_k(x)}
\end{aligned}
$$

as $n\to\infty$. It follows that (18.3) holds with $\alpha=0$ and $p(x)$ replaced by $p_k(x)$, provided our approximations are justified. In particular, we need $\lim_{n\to\infty} n\bar{F}_\varepsilon(u/k+b_n) = e^{-u}$ and $\lim_{n\to\infty} n\times\eta_k\left(\frac{u}{k}+b_n\right)=0$. The first of these two equations follows from

$$
\begin{aligned}
\lim_{n\to\infty} n\bar{F}_\varepsilon(u/k+b_n) &= \lim_{n\to\infty} n\bar{F}_\varepsilon(b_n)\frac{\sum_{m=0}^{k-1}e^{-kb_n}k^m(u/k+b_n)^m/m!}{\sum_{m=0}^{k-1}e^{-kb_n}k^mb_n^m/m!}e^{-u} \\
&= e^{-u}
\end{aligned}
$$

as $\bar{F}_\varepsilon(b_n)=1/n$, $b_n\to\infty$ and u/k is bounded, and for the second equation we use that

$$
\begin{aligned}
\lim_{n\to\infty} n\eta_k\left(u/k+b_n\right) &= \lim_{n\to\infty}\left|n\left\{\bar{F}_\varepsilon(u/k+b_n)p_k(x)-\left(1-\mu\left(x;(-\infty,u/k+b_n]\right)\right)\right\}\right| \\
&= \lim_{n\to\infty}\left|n(\bar{F}_\varepsilon(u/k+b_n)\right|\left|\frac{\bar{F}_\varepsilon(u/k+b_n)p_k(x)-\int_{\underline{y}}^{\bar{y}}\bar{F}_\varepsilon(u/k+b_n-y)dF_{h_k(x)}(y)}{\bar{F}_\varepsilon(u/k+b_n)}\right| \\
&= 0.
\end{aligned}
$$

The first term on the second line is bounded and it can be checked that the second term converges to zero, and our result follows.

Given that Assumption 1 holds with $p(x)$ replaced by $p_k(x)$, we get the argmax measure

$$
C(A;k) = \frac{\int_A p_k(x)d\Lambda(x)}{\int_\Omega p_k(x)d\Lambda(x)}.
$$

This measure has a nice consistency property when $k \to \infty$. Indeed, by the Law of Large Numbers, as $k \to \infty$ the probability distribution of $h_k(x)$ converges to a point mass at $h(x) = E_g(x) = \int_0^1 g(s)x(s)ds$, so $p_k(x) \sim e^{kh(x)}$. We define the maximum value that h attains as

$$\bar{h} = \sup_{h'} \left\{ h' : \Lambda(x : h(x) \leq h') < 1 \right\}.$$

This definition ensures that $A_\delta = \{x \in \Omega : h(x) > \bar{h} - \delta\}$ has non-zero Λ-measure for every δ.

Now, this means that $$\lim_{k\to\infty} \frac{C(A_\delta; k)}{C(\Omega - A_{2\delta}; k)} \geq \lim_{k\to\infty} \frac{e^{k(\bar{h}-\delta)}\Lambda(A_\delta)}{e^{k(\bar{h}-2\delta)}\Lambda(\Omega - A_{2\delta})} = \infty,$$

Hence $\lim_{k\to\infty} C(A_\delta; k) = 1$ for all $\delta > 0$. We can interpret this as when k grows, the choice becomes less random from the point of view of the statistician and the agent will choose the option x with the highest expected value $h(x)$ with probability 1.

18.6 Extension

We have derived a way to calculate the asymptotic behavior of the best choice $C_n = X_{[n:n]}$, and have done so for a number of assumptions on the joint distribution of (X_i, U_i) of characteristics and values. However, in order to extend our results to a wider range of distributional assumptions, we must relax the requirement that $X_{[n:n]}$ should converge to a non-degenerate distribution. For example, when X and U are distributed bivariate normally with positive correlation, $|X_{[n:n]}| \to \infty$ almost surely, whereas for other models, $X_{[n:n]}$ converges to a one-point distribution.

In these cases, it can nevertheless be possible to find a sequence of functions h_n such that $h_n(X_{[n:n]}) \Rightarrow C$ for a non-degenerate random variable C. In this case, we would have $X_{[n:n]} \overset{d}{\approx} h_n^{-1}(C)$ for large n, where $\overset{d}{\approx}$ means that the two random variables have approximately the same distribution.

We have done some exploratory studies on this extension, and there are indications that for a much larger class of distributions than studied in the present paper, it is possible to find sequences h_n and g_n such that $\sum_{i=1}^{n} \delta_{(h_n(X_i), g_n(U_i))} \Rightarrow_p \xi$ for some non-degenerate Poisson process ξ with intensity measure μ on $E = \Omega \times \mathbb{R}$. The asymptotic argmax distribution of $h_n(X_{[n:n]})$ is then $C(A) = \int_{\mathbb{R}} \frac{\mu(A, dx)}{\mu(\Omega, dx)} F_U(dx)$, for all $A \in \mathcal{B}(\Omega)$, where $U = M_\xi(\Omega)$ is the maximum utility of ξ. In particular, if $\mu = \Lambda_p \times \nu$, this argmax distribution coincides with the one in Theorem 18.2. We also conjecture that this extension can be connected to the theory of conditional extreme values, as discussed in Heffernan and Tawn [8].

18.7 Conclusion

We have shown that point process theory can be used to derive continuous approximations of discrete choice problems with a large number of options. When the random component of utility is exponentially distributed, or a convex linear combination of beta or Pareto distributions, we have derived analytical solutions to the approximation problem. Potential applications involve commuting choices, and we have provided suggestive evidence that some observed commuting flows distributions can be justified within our framework.

However, there is still a need to generalize the theory to allow for more flexible distributional assumptions. Essentially, functional forms outside our assumed domain might lead to all choices asymptotically diverging, or asymptotically collapsing on one point. For example, if the tail of utility is too thin, the distribution of choices will converge to the set of values with the highest deterministic utility value. In other cases, $C_n(A) \to 0$ for any compact A, and the choice probabilities will drift to infinity. In Sect. 18.6, we have outlined a potential extension of that would allow for a more flexible set of assumptions on the distribution of utilities. The idea is to renormalize the characteristics space to analyze the rate at which choice probabilities converge or diverge as the number of points $n \to \infty$.

Acknowledgements The authors wish to thank Dmitrii Silvestrov for valuable comments on the manuscript. Ola Hössjer's work was supported by the Swedish Research Council, contract nr. 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine.

References

1. Anderson, S.P., De Palma, A., Thisse, J.-F.: Discrete Choice Theory of Product Differentiation. MIT Press, Cambridge (1992)
2. Ben-Akiva, M., Leman, S.: Discrete Choice Analysis: Theory and Applications to Travel Demand, vol. 9. MIT Press, Cambridge (1985)
3. Burke, M., Brown, A.L., et al.: Distances people walk for transport (2007)
4. Cox, D.R., Isham, V.: Point Processes, vol. 24. Chapman and Hall/CRC, London (1980)
5. Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society, vol. 24, pp. 180–190. Cambridge University Press, Cambridge (1928)
6. Fosgerau, M., Lindberg, P.O., Mattsson, L.-G., Weibull, J.: Invariance of the distribution of the maximum. Technical report, MPRA Paper No. 63529, University Library of Münich, Germany (2015)
7. Gnedenko, B.V.: Limit theorems for the maximum term of variational series. Dokl. Akad. Nank SSSR (1941)
8. Heffernan, J.E., Tawn, J.A.: A conditional approach for multivariate extreme values (with discussion). J. R. Stat. Soc. Ser. B **66**(3), 497–546 (2004)
9. Jacobsen, M.: Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes. Birkhäuser, Boston (2005)
10. Karpiarz, M., Fronczak, P., Fronczak, A.: International trade network: fractal properties and globalization puzzle. Phys. Rev. Lett. **113**(24), 248701 (2014)

11. Ledford, A.W., Tawn, J.A.: Concomitant tail behaviour for extremes. Adv. Appl. Probab. **30**(1), 197–215 (1998)
12. Luce, R.D.: Individual Choice Behavior a Theoretical Analysis. Wiley, New York (1959)
13. Maddison, C.J., Tarlow, D., Minka, T.: A^* sampling. Adv. Neural Inf. Process. Syst. **27**, 3086–3094 (2014)
14. Malmberg, H., Hössjer, O.: Probabilistic choice with an infinite set of options: an approach based on random sup measures. In: Silvestrov, D., Martin-Löf, A. (eds.) Modern Problems in Insurance Mathematics, pp. 291–312. Springer, Berlin (2014)
15. McFadden, D.I.: Econometric models for probabilistic choice among products. J. Bus. **53**(3), 13–29 (1980)
16. O'Brien, G.L., Torfs, P.J.J.F., Vervaat, W.: Stationary self-similar extremal processes. Probab. Theory Relat. Fields **87**(1), 97–119 (1990)
17. Resnick, S.I.: Extreme Values, Regular Variation, and Point Processes. Springer, New York (2007)
18. Resnick, S.I., Roy, R.: Random USC functions, max-stable processes and continuous choice. Ann. Appl. Probab. **1**(2), 267–292 (1991)
19. Stoev, S.A., Taqqu, M.S.: Extremal stochastic integrals: a parallel between max-stable processes and α-stable processes. Extremes **8**(4), 237–266 (2005)
20. Train, K.E.: Discrete Choice Methods with Simulation, 2nd edn. Cambridge University Press, Cambridge (2009)

Chapter 19
Nonlinear Dynamics Simulations of Microbial Ecological Processes: Model, Diagnostic Parameters of Deterministic Chaos, and Sensitivity Analysis

Boris Faybishenko, Fred Molz and Deborah Agarwal

Abstract Modeling of ecological processes is demonstrated using a newly developed nonlinear dynamics model of microbial populations, consisting of a 4-variable system of coupled ordinary differential equations. The system also includes a modified version of the Monod kinetics equation. The model is designed to simulate the temporal behavior of a microbiological system containing a nutrient, two feeding microbes and a microbe predator. Three types of modeling scenarios were numerically simulated to assess the instability caused by (a) variations of the nutrient flux into the system, with fixed initial microbial concentrations and parameters, (b) variations in initial conditions, with fixed other parameters, and (c) variations in selected parameters. A modeling framework, using the high-level statistical computing languages MATLAB and R, was developed to conduct the time series analysis in the time domain and phase space. In the time domain, the Hurst exponent, the information measure–Shannon's entropy, and the time delay of temporal oscillations of nutrient and microbe concentrations were calculated. In the phase domain, we calculated a set of diagnostic criteria of deterministic chaos: global and local embedding dimensions, correlation dimension, information dimension, and a spectrum of Lyapunov exponents. The time series data are used to plot the phase space attractors to express the dependence between the system's state parameters, i.e., microbe concentrations,

B. Faybishenko (✉)
Lawrence Berkeley National Laboratory, Energy Geosciences Division,
University of California, Berkeley, CA, USA
e-mail: bafaybishenko@lbl.gov

F. Molz
Environmental Engineering and Earth Sciences Dept, Clemson University,
Anderson, SC, USA
e-mail: fredi@clemson.edu

D. Agarwal
Lawrence Berkeley National Laboratory, Computer Research Division,
University of California, Berkeley, CA, USA
e-mail: daagarwal@lbl.gov

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_19

and pseudo-phase space attractors, in which the attractor axes are used to compare the observations from a single time series, which are separated by the time delay. Like classical Lorenz or Rossler systems of equations, which generate a deterministic chaotic behavior for a certain range of input parameters, the developed mathematical model generates a deterministic chaotic behavior for a particular set of input parameters. Even a slight variation of the system's input data might result in vastly different predictions of the temporal oscillations of the system. As the nutrient influx increases, the system exhibits a sharp transition from a steady state to deterministic chaotic to quasi-periodic and again to steady state behavior. For small changes in initial conditions, resulting attractors are bounded (contrary to that of a random system), i.e., may represent a 'sustainable state' (i.e., resilience) of the ecological system.

Keywords Nonlinear dynamics · Deterministic chaos · Microbial systems modeling · Criteria of chaos · Attractor

19.1 Introduction

Although ecological and environmental systems are often considered as either deterministic or stochastic, understanding of their complexity and high-dimensionality requires an application of a combination of physical, mathematical and computational techniques, including methods of nonlinear dynamics and deterministic chaos. It is well known from studies of nonlinear dynamical systems and chaos that even deterministic systems can appear graphically to be random. For such systems, which are called deterministic chaotic, the time series is predictable on short time scales, which then diverge over longer time scales, while the system's attractor may remain bounded and predictable. The deterministic chaotic systems are characterized by high sensitivity to initial conditions and parameters, which is a typical feature of ecological processes. These ecological processes often exhibit abrupt changes and oscillations in ecosystem quality and properties, so that the ecological drivers may generate large responses or thresholds in ecosystem behavior (e.g., [10, 28]). Analysis and predictions of terrestrial and aquatic ecological processes, vitally needed for environmental management, are complicated due to the simultaneous and competing effects of multiple factors and processes controlling nonlinear dynamics taking place at multiple spatial and temporal scales. Sources of complexity affecting predictions of ecological systems are arising in resource competition theory, epidemiological theory, environmentally driven population dynamics, climatic predictions, flow and mass transport through heterogeneous media, and many others. Other sources of complexity are associated with uncertainty of monitoring/observations due to insufficient resolution of measurements, size of monitoring sensors, which are often inconsistent with the scale of processes and the scale of interest, etc.

For the last ~50 years, progress in investigations of physical phenomena known as nonlinear dynamics or simply as chaos theory drastically improved our understanding

of complex systems dynamics in many fields such as chemistry, physics, biology, geology, hydrology, economics, medicine, neurosciences, psychology, military, and many others. The study of ecological processes is ultimately related to the concept of ecological resilience, which has been around for some 30–50 years (e.g., [19–21]). Some authors define resilience as the time it takes for a system to recover from a disturbance, which Holling [25] defines as engineering resilience. In contrast, the amount of disturbance necessary to change the state of an ecosystem is known as ecological resilience [25]. Consequently, the ecological resilience of a system can be changed by shifts in the areas surrounding that ecosystem. Sharp transitions between ecological regimes have been observed in many systems, which are important for the development of alert systems of ecological systems, and to evaluate the impact of environmental changes on ecological dynamics [9]).

Many models describing the interactions of ecological populations have been derived and used for simulations of ecological processes, mostly based on modifications of Lotka–Volterra equations (e.g., [16, 29]). These equations are mainly based on the assumption of the unlimited nutrient supply into the system and linear parameters of the competitive Lotka–Volterra equations. The well-known system of Lotka–Volterra equations, initially developed in the 1920s, was further extended to include the dynamics of natural populations of predators and prey and a functional response, i.e., the Rosenzweig–McArthur model [34]. An alternative to the Lotka–Volterra predator-prey model is the Arditi–Ginzburg model to express the common prey dependent generalization [4], and the validity of prey or ratio dependent models has been further discussed in [3]. It was shown in [23] that population transitions can take place due to variations in population density, i.e., endogenous effects, or in the environmental parameters, i.e., exogenous effects, or the effects of model parameters. The Lotka–Volterra model can be generalized to any number of species competing against each other. For example, Smale [35] showed that the Lotka–Volterra system with five or more species ($N \geq 5$) can exhibit any asymptotic behavior, including a fixed point, a limit cycle, a torus or attractors, and Hirsch [24] showed that competitive Lotka–Volterra systems could not exhibit a limit cycle for the total number of interacting species $N < 3$. The competing interaction of microbiological components (fluxes, plants, microbes, etc.) and their synergistic and/or antagonistic feedbacks lead to nonlinearity of microbial systems, and the effects of nonlinearity are hallmarks of deterministic chaos.

However, research of deterministic chaotic dynamics in biological systems has not received as much attention as that in electronic, chemical, fluid-mechanical systems, meteorological studies [37] and hydrological systems [12, 32], and mathematical models are at an early stage of development [15]. However, this has recently started to change. For example, in their reviews of the current state of the problem, Faybishenko and Molz [13] and Molz and Faybishenko [31] concluded that the papers by Becks et al. [5], Graham et al. [17], and Beninca et al. [8], using experimental studies and relevant mathematics, provided breakthrough demonstrations that deterministic chaos is present in relatively simple biochemical systems of an ecological nature.

The objectives of this chapter are: (1) to develop a coupled, nonlinear dynamics mathematical model, including trophic interactions, to describe microbial processes;

(2) to carry out simulations to assess the effects of (a) endogenous factors (i.e., the nutrient flux into the system), (b) exogenous factors (i.e., variations in initial conditions), and (c) model parameters, and (3) to conduct nonlinear dynamics analysis of simulated time series to assess transitions between steady state, deterministic chaotic and quasi-periodic modulations of the nutrient and microbial concentrations. As a motivation for model development, we used the results of experimental investigations of a nutrient and 3-microbe system performed in a chemostat [5]. The basis for the development of our mathematical analysis was a model of the nutrient and 2-microbe system formulated in [26].

The rest of the paper is structured as follows: Sect. 19.2 presents a summary of the model development motivating experiments; Sect. 19.3 includes the development of the new mathematical model; Sect. 19.4 includes a general description of simulation methods and time series data analysis in time domain and phase space; Sect. 19.5 presents the results of simulations and data analysis; and Sect. 19.6 gives conclusions and several directions for future research.

19.2 Model-Motivating Experiments

Graham et al. [17] reported experimental results demonstrating the phenomenon of chaotic instability in biological nitrification in a controlled laboratory environment. In this study, the aerobic bioreactors (aerated containers of nutrient solution and microbes) were filled initially with a mixture of wastewater from a treatment plant and simulated wastewater, involving a mixture of many microbes. The main variables recorded as a time series were total bacteria, ammonia-oxidizing bacteria (AOB), nitrite-oxidizing bacteria (NOB), and protozoa, along with effluent concentrations of nitrate, nitrite and total ammonia. The method of Rosenstein et al. [33] was used to calculate maximum Lyapunov exponents, which fell roughly in a range from 0.05 to 0.2 d^{-1}. Graham et al. [17] concluded that nitrification is prone to chaotic behavior because of a fragile AOB-NOB mutualism, i.e., interaction.

Beninca et al. [8] conducted a laboratory experiment over a period of 6.3 years, which demonstrated chaotic dynamics in a plankton community in a water sample obtained from the Baltic Sea. This experiment was housed in a cylindrical container that was 45 cm in diameter, 74 cm high and filled with 90 l of water with a 10 cm sediment layer at the bottom. The predictability of each data set was shown to decrease with time (essentially lost after 15–30 days), consistent with positive Lyapunov exponents averaging about 0.058 per day. They were calculated based on 2 methods: attractor reconstruction using time-delay plots, and direct calculation of the Lyapunov exponents [22]. It is apparent that when dealing with exponential divergence of trajectories in phase-space, the dominant and positive Lyapunov exponent does not have to be large for the phenomenon of chaotic dynamics to occur over time scales of interest.

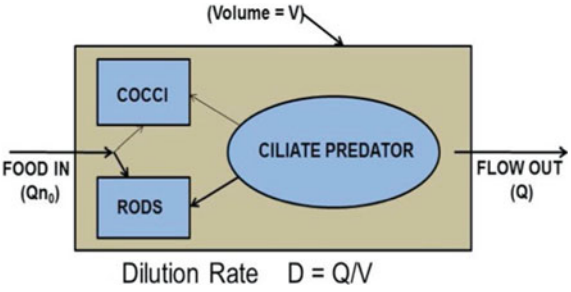

Fig. 19.1 Schematic of Becks et al. experiments (modified from [5])

In their particularly detailed experiments, Becks et al. [5–7] studied a microbial food web in a chemostat. A conceptual model of these experiments is shown in Fig. 19.1.

The food web was composed of a nutrient source, two bacteria that consumed nutrient (a rod and a coccus), and a ciliate predator that consumed both bacteria. The variable driving the system was the food supply that was varied by changing the dilution rate. The four coupled dependent variables were concentrations of nutrient (mg/cc) and each of the three microbes (cells/cc). For a fixed set of dilution rates, the three microbe concentrations were measured at a selected set of approximately daily time intervals. (Although Becks et al. [5] expressed the time units of days, in this chapter, we provided time units in both days and hours). Each set of data constituted a time series of concentrations at discrete times, and deterministic chaos was identified by using a computerized version of the analytical procedure developed by Rosenstein et al. [33] for calculating the largest Lyapunov exponent. In the chaotic data range, which averaged a dilution rate of about $0.5\,d^{-1}$, the dominating Lyapunov exponents had statistically significant values of about $0.18\,d^{-1}$. The authors concluded that classical steady states were observed at $D = 0.75\,d^{-1}$ and $D = 0.9\,d^{-1}$, chaotic dynamics were observed at $D = 0.5\,d^{-1}$, and periodic dynamics were observed at a slightly lower $D = 0.45\,d^{-1}$. The chaotic dynamics observed by Becks et al. [5] took place under conditions of a constant feeding rate. Therefore, the observed chaotic dynamics were internal to the system, which may be a type of emergence in phase space.

19.3 Mathematical Model Development

The model used in the development of our new mathematical model is a 3-equation model (nutrient plus 2 microbes) developed by Kot et al. [26]. This model is generalized to a 4-variable coupled system of equations to describe nutrient concentration $n(t)$, rod population $r(t)$, cocci population $c(t)$ and predator population $p(t)$. The units of the nutrient concentration were mg/cc, while those of the microbe populations were cell numbers/cc (cells/cc). Thus in order to conserve mass in the resulting

model, an average microbial mass had to be selected for each microbe type. This was done based on measured microbe volume averages published by Becks et al. [5], along with an assumed density of 1 g/cc: r mass $m_r = 1.6\text{E-}9$ mg; c mass $m_c = 8.2\text{E-}9$ mg, and p mass $m_p = 3.2\text{E-}6$ mg. With these minor differences, the Kot et al. [26] Eq. (19.1) for a nutrient (mg/cc), a rod (cells/cc), and a predator (cells/cc) would be written as:

$$\begin{aligned}
\frac{dn}{dt} &= Dn_o - \frac{\mu_{rn}}{Y_{rn}}\left[\frac{n(rm_r)}{K_{rn}+n}\right] - Dn \\
\frac{d(rm_r)}{dt} &= \mu_{rn}\left[\frac{n(rm_r)}{K_{rn}+n}\right] - \frac{\mu_{pr}}{Y_{pr}}\left[\frac{(rm_r)(pm_p)}{K_{pr}+(rm_r)}\right] - D(rm_r) \\
\frac{d(pm_p)}{dt} &= \mu_{pr}\left[\frac{(rm_r)(pm_p)}{K_{pr}+rm_r}\right] - D(pm_p)
\end{aligned} \tag{19.1}$$

These three equations are identical to those used by Kot et al. [26]. (The Kot et al. notations S, H, and P are being changed in Eqs. (19.1) to n, rm_r and pm_p, respectively.) All the m_x terms are mean mass per respective microbe, so r and p are dimensionless numbers of rods and predators per cc, the microbe concentration units recorded by Becks et al. [5]. All the other constant terms are maximum specific growth rates μ_{xx}, half saturation constants K_{xx} and yield coefficients Y_{xx}.

The system of Equations (19.1) for 3 microbes can be extended to include a 2nd nutrient-consuming microbe, a cocci, by adding an equation similar to the 2nd equation of this system, along with the analogous coupling terms, resulting in the following system of 4 ordinary differential equations:

$$\begin{aligned}
\frac{dn}{dt} &= Dn_o - \frac{\mu_{rn}}{Y_{rn}}\left[\frac{n(rm_r)}{K_{rn}+n}\right] - \frac{\mu_{cn}}{Y_{cn}}\left[\frac{n(cm_c)}{K_{cn}+n}\right] - Dn \\
\frac{d(rm_r)}{dt} &= \mu_{rn}\left[\frac{n(rm_r)}{K_{rn}+n}\right] - \frac{\mu_{pr}}{Y_{pr}}\left[\frac{(rm_r)(pm_p)}{K_{pr}+(rm_r)}\right] - D(rm_r) \\
\frac{d(cm_c)}{dt} &= \mu_{cn}\left[\frac{n(cm_c)}{K_{cn}+n}\right] - \frac{\mu_{pc}}{Y_{pc}}\left[\frac{(cm_c)(pm_p)}{K_{pc}+(cm_c)}\right] - D(cm_c) \\
\frac{d(pm_p)}{dt} &= \mu_{pr}\left[\frac{(rm_r)(pm_p)}{K_{pr}+rm_r}\right] + \mu_{pc}\left[\frac{(cm_c)(pm_p)}{K_{pc}+cm_c}\right] - D(pm_p)
\end{aligned} \tag{19.2}$$

The parameters involved are: maximum specific growth rates (μ_{xx}), half saturation constants (K_{xx}) and yield coefficients (Y_{xx}), a total of twelve. Equations (19.2) are direct generalization of the Kot et al. model [26], and their mathematical validity was checked by showing that a mass balance was maintained, and when the Kot et al. [26] initial conditions and parameter values were used, with one microbe dying out, the results of Kot et al. ([26] Figs. 3 and 6) were reproduced.

However, the system of Equations (19.2) do not include a term describing the process of nutrient recycling, which could have occurred in the Becks et al. [5] experiments. The specific death rate of predators and their biomass recycling to nutrient are expected to be important, because their unit mass is about 1000 times greater than that of each feeding microbe. Moreover, populations of feeding microbes decrease mainly due to consumption by predators. When predators die, their bodies simply break down with remains consumed or flushed out of the system. Natural death and biomass recycling of the feeding microbes is assumed to be negligible relative to predators, and this was supported by numerical experiments. The specific death rate for predators and the nutrient recycling terms will be identified in the final set of Equations (19.10).

As observed in supplemental experiments by Becks et al. [5], in the absence of predators, r was able to out-compete c for nutrient, and at an identical population of r and c (4E6 cells/cc), p consumed r over c in the ratio of 4:1. In the absence of further guidance from the experiments, we decided to incorporate this additional information into Eq. (19.2) as follows: (a) When r and c are low, p chooses them on an equal basis even though in general r is preferred over c (assuming that starving organisms are not choosy), (b) at high r and c, as observed in the supplemental experiments, p chooses r four times more than c, and (c) in competition for nutrient with no predators, the cocci would die out first.

A simple way to impose the condition (c) is to set the value of μ_{rn}, the maximum specific growth rate of r on n, equal to $k\mu_{cn}$, with $k > 1$. Then for k sufficiently large, r will always outcompete c. Incorporating conditions (a) and (b) is more involved, but it is still straightforward as described below.

Based on the Becks et al. ([5], Fig.1) data, the r and c concentrations are ranging from about 1E5 cells/cc to 2E6 cells/cc. According to the 2nd and 3rd equations of the system of Equations (19.2), the respective uptake rates of p on r, expressed as $\mathrm{d}r_p/\mathrm{d}t$, and p on c, expressed as $\mathrm{d}c_p/\mathrm{d}t$, are given by

$$\begin{aligned} \frac{dr_p}{dt} &= \frac{pm_p\mu_{pr}r}{Y_{pr}(K_{pr}+m_r r)} \\ \frac{dc_p}{dt} &= \frac{pm_p\mu_{pc}c}{Y_{pc}(K_{pc}+m_c c)} \end{aligned} \tag{19.3}$$

For the minimum values of r and c ($\sim$1E5 cells/cc), assuming $dr_p/dt = dc_p/dt$, the system (19.3) yields

$$\frac{pm_p\mu_{pr}(1E5)}{Y_{pr}(K_{pr}+1E5m_r)} = \frac{pm_p\mu_{pc}(1E5)}{Y_{pc}(K_{pc}+1E5m_c)} \tag{19.4}$$

Assuming that for maximum values of r and c (~2E6 cells/cc), $dr_p/dt = 4dc_p/dt$, we obtain:

$$\frac{pm_p\mu_{pr}(2E6)}{Y_{pr}(K_{pr}+2E6m_r)} = \frac{4pm_p\mu_{pc}(2E6)}{Y_{pc}(K_{pc}+2E6m_c)} \tag{19.5}$$

To achieve the equalities specified in Eqs. (19.4) and (19.5), the numerator and denominator of dr_p/dt will need to be modified, so that the maximum specific growth rate of p on r will become $\mu_{pr}(m_1 r + i_1)$, and the denominator will become $(K_{pr} + m_2 r + m_r r)$, with parameters m_1, i_1 and m_2. Thus, Eq. (19.3) for dr_p/dt yields

$$\frac{dr_p}{dt} = \frac{pm_p\mu_{pr}(m_1 r + i_1)r}{Y_{pr}(K_{pr}+m_2 r + m_r r)} \tag{19.6}$$

In Eq. (19.6), m_1 is in units of $(cells/cc)^{-1}$, m_2 is in units of mg, and i_1 is dimensionless. Equation (19.6) is a new semi-empirical relationship expressing a modified form of the Monod kinetics equation, partly motivated by experiment, to take into account a p preference change for r relative to c. Then to satisfy Equations (19.4) and (19.5), the following conditions must be met:

$$\begin{aligned} \frac{\mu_{pr}(1E5m_1 + i_1)}{Y_{pr}(K_{pr}+1E5m_2+1E5m_r)} &= \frac{\mu_{pc}}{Y_{pc}(K_{pc}+1E5m_c)} \\ \frac{\mu_{pr}(2E6m_1 + i_1)}{Y_{pr}(K_{pr}+2E6m_2+2E6m_r)} &= \frac{4\mu_{pc}}{Y_{pc}(K_{pc}+2E6m_c)} \end{aligned} \tag{19.7}$$

The introduced parameters m_1 and i_1 must satisfy the conditions: $1E5m_1 + i_1 = 1$, and $2E6m_1 + i_1 = 4$, also setting $\mu_{pr} = \mu_{pc}$. These conditions yield $m_1 = 1.579E-6(\text{cells/cc})^{-1}$, $i_1 = 0.8421$. The corresponding conditions on m_2 are:

$$\begin{aligned} Y_{pr}(K_{pr}+1E5m_2+1E5m_r) &= Y_{pc}(K_{pc}+1E5m_c) \\ Y_{pr}(K_{pr}+2E6m_2+2E6m_r) &= Y_{pc}(K_{pc}+2E6m_c) \end{aligned} \tag{19.8}$$

If we set $Y_{pr} = Y_{pc}$ and $K_{pr} = K_{pc}$, Eq. (19.8) become:

$$1E5m_2 + 1E5m_r = 1E5m_c, \quad 2E6m_2 + 2E6m_r = 2E6m_c \tag{19.9}$$

It can be seen from both relationships that $m_2 = m_c – m_r$ = 8.2E-9 – 1.6E-9 = 6.6E-9 mg. So, Eq. (19.2) adapted to the Becks et al. supplemental experiments after dividing through by the microbial masses may be written as follows:

$$\begin{aligned}
\frac{dn}{dt} &= Dn_o - \frac{\mu_{rn}}{Y_{rn}}\left[\frac{n(rm_r)}{K_{rn}+n}\right] - \frac{\mu_{cn}}{Y_{cn}}\left[\frac{n(cm_c)}{K_{cn}+n}\right] - Dn + pm_p\delta_p(EF)\\
\frac{dr}{dt} &= \mu_{rn}\left[\frac{nr}{K_{rn}+n}\right] - \frac{\mu_{pr}(1.58\text{E-}6r + 0.842)}{Y_{pr}}\left[\frac{r(pm_p)}{K_{pr} + 6.6\text{E-}9r + rm_r}\right] - Dr\\
\frac{dc}{dt} &= \mu_{cn}\left[\frac{nc}{K_{cn}+n}\right] - \frac{\mu_{pc}}{Y_{pc}}\left[\frac{c(pm_p)}{K_{pc}+(cm_c)}\right] - Dc\\
\frac{dp}{dt} &= \mu_{pr}(1.58\text{E-}6r + 0.842)\left[\frac{p(rm_r)}{K_{pr} + 6.6\text{E-}9r + rm_r}\right]\\
&+ \mu_{pc}\left[\frac{p(cm_c)}{K_{pc}+cm_c}\right] - Dp - p\delta_p
\end{aligned} \tag{19.10}$$

In these equations, δ_p is the specific death rate for predators, which is recycled to nutrient at the efficiency $EF \leq 1$ in the 1st equation of the set (19.10). Equations (19.10) are also subject to the parameter restraints, resulting from conditions (19.7), given by:

$$\begin{aligned}
\mu_{rn} &= k\mu_{cn}, \quad \mu_{pr} = \mu_{pc}\\
K_{pr} &= K_{pc}, \quad Y_{pr} = Y_{pc}
\end{aligned} \tag{19.11}$$

Because the predators have been made to prefer rods over cocci, the rods are disadvantaged and would tend to die out. Based on numerical experiments, this is prevented by setting the k factor in the first of Equations (19.11) to 1.5.

19.4 Model Parameters Used for Simulations

Three types of modeling scenarios were developed and applied to assess the transition from steady state to quasi-periodic and to deterministic chaotic behavior, with (a) varying nutrient flux (i.e., endogenous scenarios), (b) varying initial concentrations (i.e., exogenous scenarios), and (c) varying input model parameters, while other parameters were fixed.

19.4.1 Scenarios A: Varying Nutrient Flux, with Fixed Initial Concentrations and Model Parameters

The system of Equations (19.10) was solved subject to a range of initial conditions and parameters, which were selected in the range of values observed in Becks et al. experiments [5], as well as with some deviations to assess the performance of the system of Equations (19.10). In these scenarios, the dilution rate D ranged from 0.35 d^{-1} to 5 d^{-1}, with the following initial concentrations at t = 0: n = 0.085 mg/cc,

Table 19.1 Model parameters used for simulations. (Note in Table 19.1: Left column indicates subscripts used in Eqs. (19.1)–(19.11), and the upper row indicates the variables used in these formulae.)

Variables	$\mu(h^{-1}/d^{-1})$	K (mg/cc)	Y	m
rn	0.1873/4.4952	0.009	0.4	
cn	0.1248/2.9952	0.009	0.4	
pr	0.05117/1.228	0.009	0.6	
pc	0.05117/1.228	0.009	0.6	
r				1.6E-9
c				8.2E-9
p				3.2E-6

r = 4.2E6 cells/cc, c = 1E6 cells/cc, and p = 3E3 cells/cc, with n_0 held constant at 0.15 mg/cc, EF = 0.5, δ_p = 0.0998 1/d. The other model parameters are listed in Table 19.1.

19.4.2 Scenarios B: Varying Initial Microbial Concentrations, with Fixed Nutrient Flux and Model Parameters

The initial concentrations were assigned in the following ranges: r from 2.1E5 to 4.263E6 cells/cc, c from 5E5 to 1.015E6 cells/cc. Simulations were conducted for the fixed dilution rate of $D = 0.4995\, d^{-1}$, and a slightly increased (compared to Scenarios A) fixed initial nutrient concentration of $n = 0.103$ mg/cc.

19.4.3 Scenarios C: Fixed Nutrient Flux and Initial Conditions, with Varying Model Parameters

In these simulation runs, we changed the values of model parameters of the modified Monod kinetics equation (Eq. (19.6)) m_1 and m_2. Note that Eq. (19.6) is a new semi-empirical relationship expressing a modified form of the Monod kinetics equation, which takes into account a preference change for r relative to c. In simulations, m_1 changed from 1.6E-06 to 2.24E-06 (cells/cc)$^{-1}$, and m_2 changed from 3.96E-09 to 5.94E-09 mg. The other parameters used in Scenarios C remained as those given in Scenarios A.

19.5 Methods of Simulations and Data Analysis

19.5.1 Method of Simulations

The system of 4-coupled ordinary differential equations, Eq. (19.10), was solved using MATLAB software with the application of the solver ODE45. The ODE45 solver implements a Runge–Kutta (RK) algorithm [11]. This algorithm uses a combination of 4th order and 5th order RK formulae to solve the ODE system and make time-step adjustments to control truncation error. Total accumulated error is of order Δt^4. The length of run times was 8,000 h (333 d), and the time steps varied from 0.28 to 8.1 h, with the average value of 2.72 h. In order to apply the methods of the time series analysis described in Sect. 19.5.2, the simulated irregular time series was converted to a regular time series with time steps of 0.5 h using the spline function of MATLAB.

19.5.2 Methods of Time Series Analysis

19.5.2.1 Time Series Analysis in the Time Domain

The detrended fluctuation analysis (DFA) is applied to determine the statistical self-affinity of a time series signal by means of calculating the scaling exponent or Hurst exponent. This type of analysis is used to characterize the long memory dependence of the time series. A system is self-affine if the following relation holds [36, 37]:

$$\Delta x(\lambda \Delta t) = \lambda^H \Delta x(\Delta t)$$

where H is the Hurst exponent, i.e., the scaling exponent, $0 < H < 1$. For $H \to 0$, the position at any time is independent of the position at any previous time, and the process may correspond to white noise. As H increases, the dependence of the position at a given moment of time on the position in the past increases, implying better predictability of the system behavior. For a quasi-periodic system, when fluctuations display long-range correlations of a dynamical system far from equilibrium, the Hurst exponent is approaching 1, and for a time series with severe instabilities, the DFA shows a breakdown of the long-range correlation behavior, and the Hurst exponent is decreasing. The Hurst exponent is a characteristic of the "fractality," or scaling, in time series.

Correlation (delay) time, τ, is the time between discrete time-series points, when the relationship between the points practically vanishes. It was determined using the average mutual information function. This function provides the information learned about one observation from another one on the average over all measurements [1]. The mutual information function is always $I(\tau) \geq 0$. This function determines the amount of information, in bits, learned about the point $P(t + n\tau)$, at the time

$(t + n\tau)$, from the measurements of $P(t + (n - 1)\tau)$, at the time $(t + (n - 1)\tau)$. The first minimum of the average mutual information function versus τ is used to determine an optimum correlation time. The selection of τ within a small range around the first minimum of the mutual information function will not affect the results of calculations for diagnostic parameters of chaos described below [1].

19.5.3 Time Series Analysis in the Phase Space Domain

Global embedding dimension, D_{GED}, is the dimension of the phase space needed to unfold the attractor of a nonlinear system from the observed scalar signals in such a way that the trajectories of the D_{GED}-dimensional attractor no longer cross each other. The global embedding dimension is determined using a method called false nearest neighbors (FNN) [1], which is used to determine whether a nearest neighbor of a certain point in phase space is a true neighbor characterizing a dynamic system or a false projection caused by a too low dimensional phase space [1, 2]. For a purely deterministic chaotic system, such as the Lorenz model, the percentage of false nearest points drops to zero at $D_{GED} = 3$, and remains zero for higher dimensions, implying that the attractor is fully unfolded and remains unfolded as the embedding dimension increases. The value of D_{GED} corresponds to the minimum value of FNN. In the presence of a random component (noise), the FNN curves do not drop to zero, and the minimum value of the FNN corresponds to the embedding dimension that is slightly higher than that for a clean (not noisy) data set.

The *local embedding dimension*, D_L, characterizes how the dynamic system evolves on a local scale. The D_L indicates the number of degrees of freedom governing the system dynamics, i.e., how many dimensions should be used to predict the system dynamics [1, 2]. The local embedding dimension, D_L, is less than or equal to the global embedding dimension, D_{GED}.

The *correlation dimension*, D_{cor} (also called D_2) is a scaling exponent characterizing a cloud of points in an n-dimensional phase space, and is calculated from [18] as $C(r) \sim r^{D_{cor}}$ or as the slope of $\log(C(r))$ versus $\log r$: $D_{cor} = d\log C(r)/d\log r$.

Lyapunov exponents are the most valuable diagnostic parameters of a deterministic chaotic system. The Lyapunov exponents are used to assess the sensitivity of a nonlinear dynamical system to small changes in initial conditions as well as truncation error and round-off error at the end of each time step in a numerical solution. All of this affects the stability and asymptotic behavior of nonstationary solutions of ODEs. The spectrum of Lyapunov exponents is determined by measuring the mean exponential growth or shrinking of perturbations with respect to a nominal trajectory. Chaotic systems have at least one positive Lyapunov exponent, and the sum of the spectrum of Lyapunov exponents is negative. Generally, a Lyapunov exponent indicates how two nearby points in a phase space of a dissipative chaotic system move exponentially apart in time. The exponentially rapid separation of initially close points leads to sensitivity of the system to initial conditions. The number of Lyapunov exponents is equal to the local embedding dimension D_L. Each Lyapunov

exponent gives the rate of expansion or contraction along the coordinates of the phase space. For a chaotic system, the largest Lyapunov exponent must be positive, which implies that a line segment in the phase space grows as $exp(\lambda_i t)$, while a negative sum of the spectrum of Lyapunov exponents ($\sum \lambda_i < 0$) indicates that the phase-space attractor does not expand indefinitely, but occupies a limited phase-space volume.

Information dimension (D_{inf}), which is a measure of how the average Shannon information scales, was computed for different embedding dimensions. As with calculations of D_{cor}, this approach was used for checking how D_{inf} saturates with the increase in the embedding dimension: if the slopes converge at the embedding dimension D_{GED}, then D_{GED} is the correct value, and the convergent value of the slope is taken as a D_{inf} estimate.

19.6 Modeling Results

19.6.1 Scenarios A

19.6.1.1 Time Domain Analysis

Based on a visual examination of the simulated time series and the following phase space analysis, we divided the time series patterns into 5 groups (a complete set of time series graphs is included in SI Appendix A). In plotting these series, the concentrations were normalized using the following normalization coefficients – for r: 4.2E6, for c: 5E6, and p: 1E3):

- Group 1: D ranges from $0.35\,d^{-1}$ to $0.4879\ d^{-1}$, exhibiting initial oscillations, which then converge to a stable equilibrium (steady state),
- Group 2: D ranges from $0.488\,d^{-1}$ to $0.489\,d^{-1}$, exhibiting deterministic chaotic oscillations,
- Group 3: D ranges from $0.490\,d^{-1}$ to $0.510\,d^{-1}$, with quasi-periodic oscillations of the nutrients, preys, and the predator,
- Group 4: D ranges from $0.515\,d^{-1}$ to $0.675\,d^{-1}$, exhibiting quasi-periodic oscillations of n, p, with dying r and growing c, and
- Group 5: $D \geq 0.7\,d^{-1}$, exhibiting various forms of converging to the equilibrium. (A complete set of time series plots is included in SI Appendix A.[1])

The DFA analysis showed that the temporal patterns and ranges of the Hurst exponent varied differently for time series of n, r, c, and p as a function of D (Fig. 19.2). For example, the Hurst exponent dropped from 1 to 0.95 for n and c, and from 1 to 0.98 for p, indicating a small amount of a random component in simulated time series, but H dropped much more significantly–to 0.6–for the r time series. The phase space analysis described below is consistent with these results.

[1] Supplementary Information containing 4 Appendices can be found at URL https://goo.gl/xdCtHK

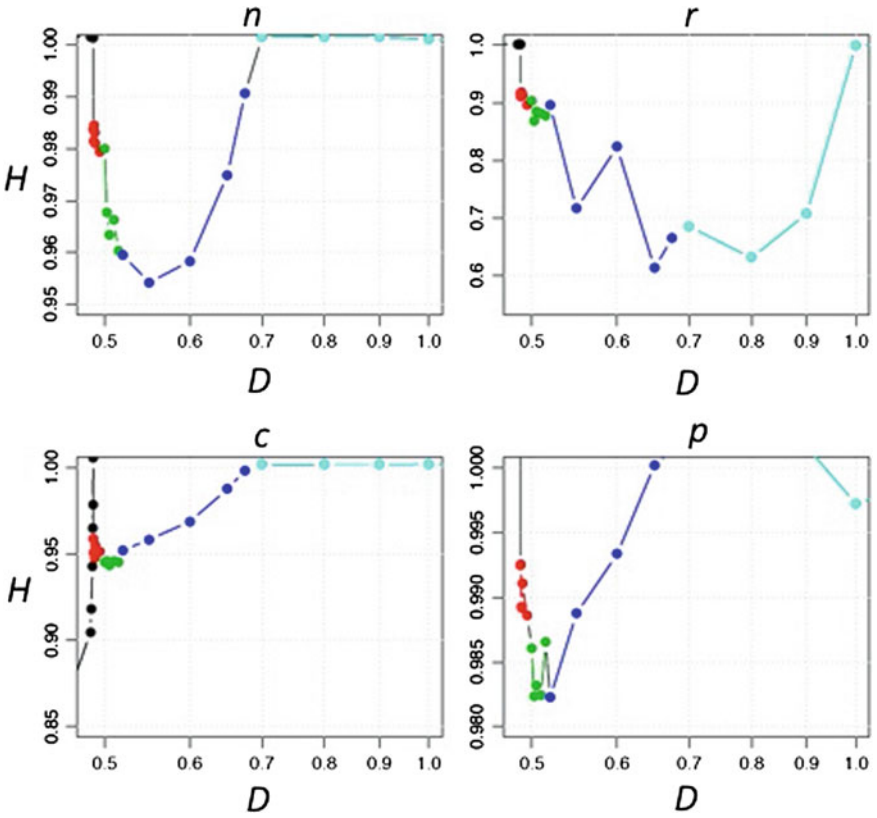

Fig. 19.2 Hurst exponent versus dilution rate: Black points – Group 1, Red points – Group 2, Green points – Group 3, Blue points – Group 4, and Light blue points – Group 5

The calculated patterns of Shannon's entropy are generally similar to those of the function H vs D, and the correlation between H and the Shannon's entropy is shown in Fig. 19.3. Note also from Fig. 19.3 that the entropy fluctuated only slightly for n (from 9.24 to 9.68) and p (from 9.54 to 9.68), and much more significantly for r (from 4.68 to 9.68) and c (from 4.56 to 9.68). Table A-1 in SI Appendix A includes statistical parameters of the calculated Shannon's entropy.

Figure 19.4 depicts the time series of concentrations of n, r, c, and p for different nutrient inflow rates, demonstrating that for D from $0.35\,d^{-1}$ to $0.4879\,d^{-1}$, after the initial period of oscillations, the system approaches a stable equilibrium (steady state). For this range of D, all concentrations are converging to stable values, with the cocci dying and rods increasing, as it was observed in Becks et al. [5] experiments. When the value of D increased slightly (only 0.02%) to 0.488, the pattern of concentrations became oscillating, as shown in Fig. 19.4 by a black line.

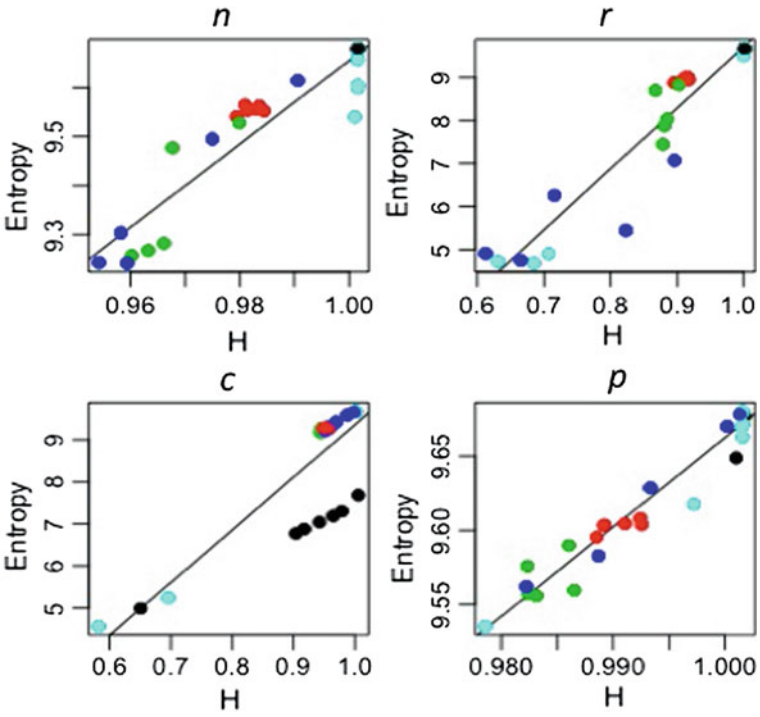

Fig. 19.3 Relationship between the Hurst exponent and Shannon's entropy

Figure 19.5 illustrates an initial period of time series oscillations, showing that the initial oscillations are practically the same for all nutrient inflow rates, indicating that for the initial period the temporal oscillations of concentrations are not sensitive to variations of D ranging from $0.35\,d^{-1}$ to $0.488\,d^{-1}$.

Table 19.2 presents the results of time lag calculations, which were then used for calculations of other phase domain parameters. Note the lower values of the time lags for n and p, which is consistent with lower Global Embedding Dimensions for these variables (see below a discussion and figures on the Global Embedding Dimension).

19.6.1.2 Phase Space Time Series Analysis

Figure 19.6 illustrates two different types of 3-D phase space attractors: the left figures demonstrate the attractors converging to a single point for $D = 0.4879\,d^{-1}$, representing Group 1; and the right figures demonstrates so called strange attractors, which are typical for a deterministic chaotic system, for $D = 0.488\,d^{-1}$, representing Group 2. The figure shows the attractors in coordinates (n, p, r), (n, p, c), (n, r, c),

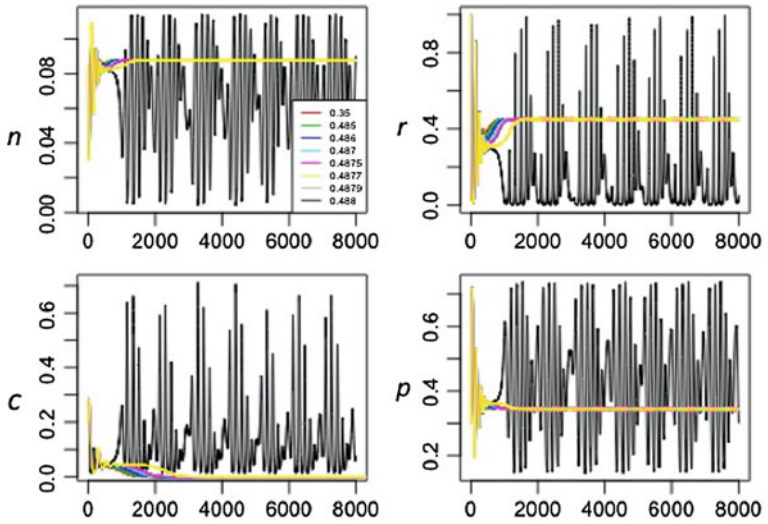

Fig. 19.4 Time series for different dilution rates, demonstrating that when D changed slightly from 0.4879 d^{-1} (color lines) to 0.488 d^{-1} (black line) concentrations began oscillating. (Nutrient inflow rates are proportional to D, i.e., Dn_o.)

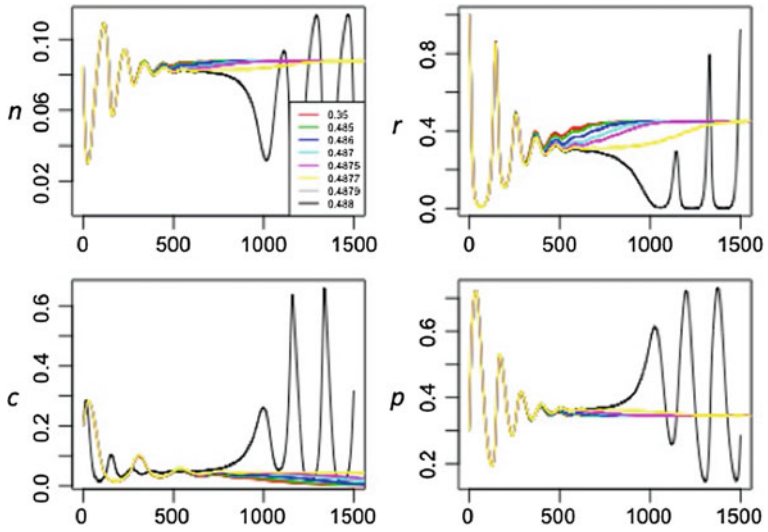

Fig. 19.5 Time series graphs showing that the initial oscillations are quite similar for all dilution rates

and (p, r, c). The phase-space strange chaotic 3-D attractors for Group 2 are shown in Fig. 19.7.

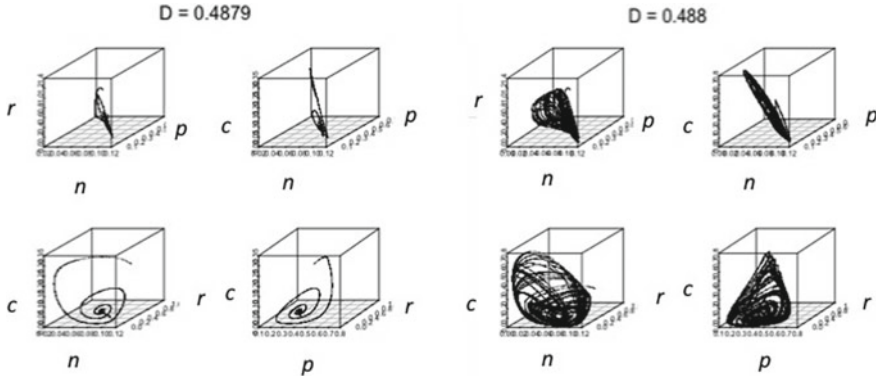

Fig. 19.6 Comparison of 3–D phase-space attractors for $D = 0.4879\ d^{-1}$ (Group 1) and $D = 0.488\ d^{-1}$ (Group 2)

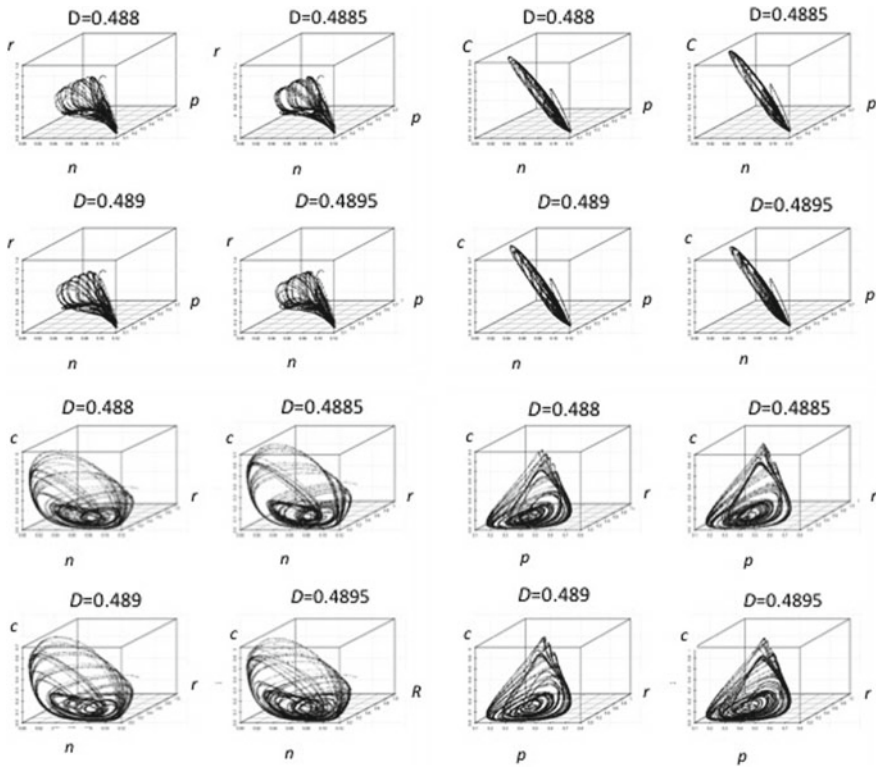

Fig. 19.7 3-D phase-space attractors for Group 2 demonstrating strange deterministic chaotic attractors

Table 19.2 Calculated time lags for different values of the dilution rate for Group 2 time series exhibiting deterministic chaotic behavior. Note: the time lags are given in units of $\frac{1}{2}$ h. The graphs of the mutual information used for calculations of time lags are given in SI Appendix A

$D(d^{-1})$	N	R	C	P
0.488	71	90	91	59
0.4885	65	87	76	59
0.489	71	95	87	59
0.4895	70	97	80	59
0.49	52	82	80	64
0.495	71	95	86	73

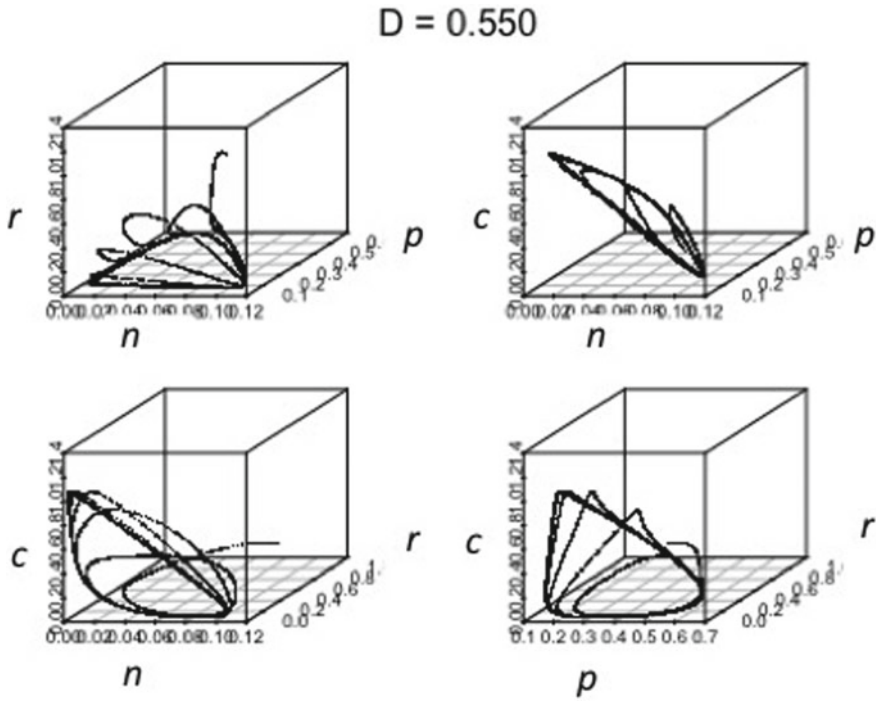

Fig. 19.8 3-D phase-space attractors for Group 3 demonstrating a quasi-periodic structure of attractors for $D = 0.550$.

As the dilution rate increased, the time series became quasi-periodic (see SI Appendix A), and the shape of the 3D attractors changed, which is demonstrated in Figure 19.8 for $D = 0.55\,d^{-1}$ (Group 3). Further increase in the nutrient influx lead to a series of converging attractors shown in Fig. 19.9 – from a simple line for $D = 0.7\,d^{-1}$ to more complex for $D = 2d^{-1}$ and $D = 3\,d^{-1}$.

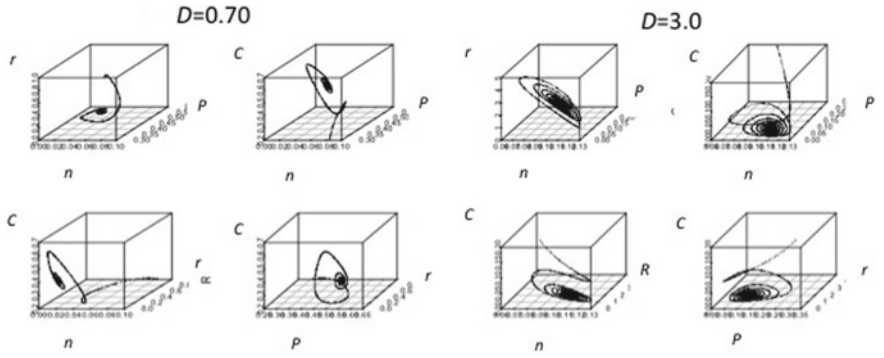

Fig. 19.9 3-D phase-space attractors for Group 4 demonstrating a converging structure of attractors

Table 19.3 Calculated correlation dimensions for Group 2

$D(d^{-1})$	n	r	c	p
0.488	2.19	2.04	2.02	2.18
0.4885	2.28	2.29	1.75	2.31
0.489	1.98	1.91	2.02	1.99
0.4895	2.12	2.13	2.13	2.15

Table 19.4 Calculated information dimension (D_{inf}) for the time series of Group 2, which are typical for deterministic chaotic behavior

$D(d^{-1})$	n	r	c	p
0.488	2.876	2.763	2.541	3.011
0.4885	2.781	2.797	2.563	2.859
0.489	2.738	2.693	2.566	2.792
0.4895	2.696	2.617	2.642	2.710
0.49	2.676	2.484	2.535	2.780
0.495	2.565	2.696	2.520	2.571

The time lags calculations were used for the evaluation of the Global Embedding Dimension (D_{GED}). The resulting plots of the FNN for n and p, which are shown in SI Appendix B, indicate that the FNN function dropped to 0 at the embedding dimension of 3, and remained 0 as the embedding dimension increased. However, for r and c the FNN functions drop significantly at $D_{GED} = 3$, but did not reach 0, followed by a gradual drop until $D_{GED} = 7$ for r and $D_{GED} = 4$–5 for c.

The values of the embedding dimension and time lags were used for calculation of the correlation dimension (D_{cor}) and information dimension (D_{inf}). Figure A-6

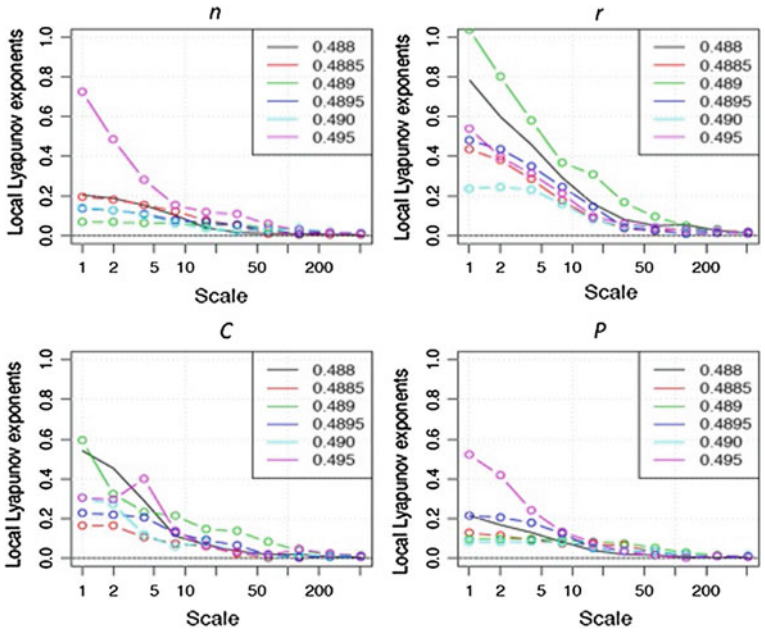

Fig. 19.10 Maximum Lyapunov exponents remain positive as the scale increases

in SI Appendix A shows the plots of the correlation dimension versus embedding dimension, and Table 19.3 provides a summary of calculated correlation dimensions.

Table 19.4 includes the results of calculations of the information dimension (D_{inf}). The values for D_{cor} and D_{inf} are typical for a deterministic chaotic behavior.

Calculations of the spectrum of Lyapunov exponents for all time series of Group 2 were conducted for the local embedding dimension of 3, i.e., 3 Lyapunov exponents were calculated for each time series. The results of calculations are plotted in SI Appendix C. All graphs in SI Appendix C show that there is one Lyapunov exponent equal to 0. Figure 19.10 shows that for all time series the largest Lyapunov exponent vs scale remained positive, and Fig. 19.11 shows that the sum of Lyapunov exponents remained negative versus scale, and that for each time series there is one Lyapunov exponent equal to 0, which are all typical features of deterministic chaos.

The Poincare maps plotted using the extrema of time series are shown in SI Appendix D. These maps show a definite structure, which is typical for a deterministic chaotic system.

Finally, using the time lags for each of the variables n, r, c, and p, given in Table 19.2, we plotted 3D pseudo-phase space attractors shown in Figs. 19.12, 19.13 for one of the dilution rates $D = 0.488\, d^{-1}$. The resulting strange attractors are well-defined and bounded, representing phenomena of forced oscillations simulated using a system of four ODEs.

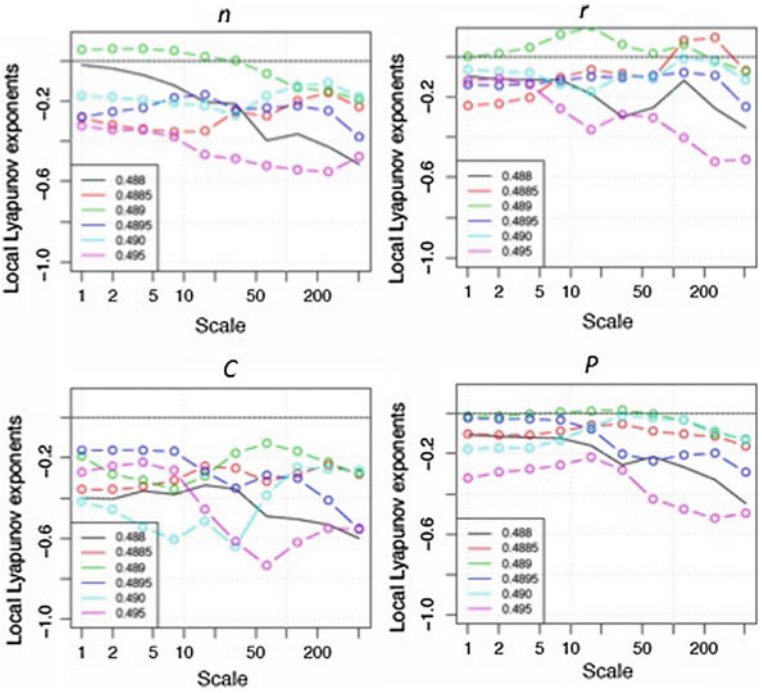

Fig. 19.11 The sum of Lyapunov exponents is negative as the scale increases, indicating the strange attractors are converging

19.6.2 Scenarios B

These calculations were conducted to assess the effects of endogenous factors, i.e., small changes in initial conditions of the system of Equations (19.10), by varying initial values of $r(t = 0)$ and $c(t = 0)$ concentrations, on the time series oscillations. Figure 19.14 shows that after an initial period of 250 h, even minute changes of the initial concentration of c lead to significant divergence of time series oscillations of all other variables – n, r, and p, and then to the system stabilization. Figure 19.15 demonstrates that minute changes in the initial concentration of r lead to changes of the time series pattern from oscillating to converging to stable conditions. Note that the reference (baseline) time series curves on Figs. 19.14 and 19.15 are shown in black color.

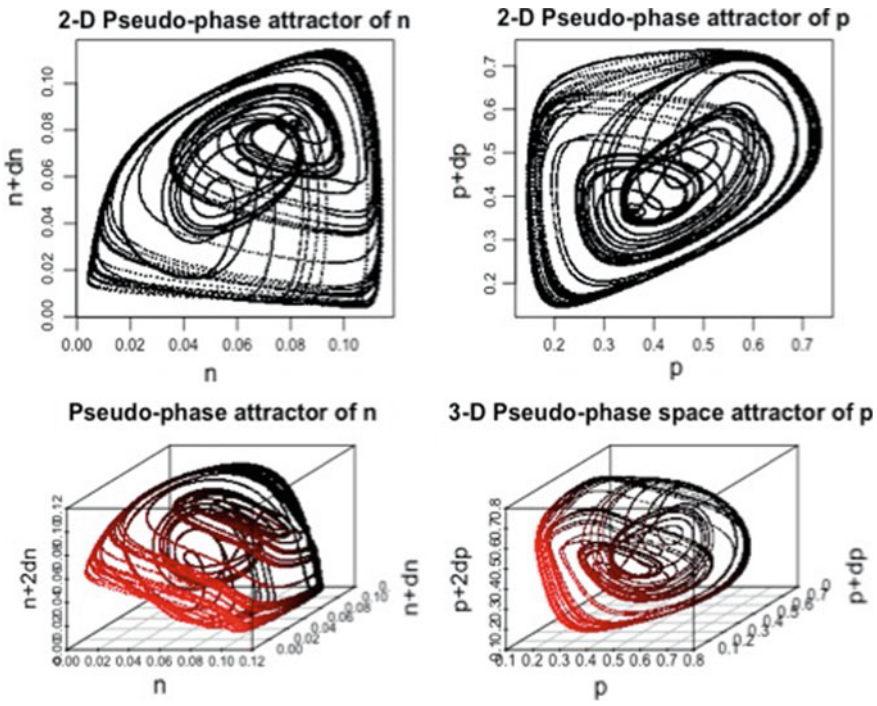

Fig. 19.12 2-D and 3-D pseudo-phase space attractors for n (left) and p (right)

19.6.3 Scenarios C

Calculations were carried out to demonstrate the effects of small changes in model parameters of m_1 and m_2 of Equations (19.10) on the results of predictions. Figure 19.16 shows that the increase in the value of m_1 by only 5% leads to the transition of oscillating behavior to the system equilibrium, with the decrease of c and increase of r.

Figure 19.17 also demonstrates an inverse relation between the concentrations of n and p, as the value of m_1 increases from 5 to 40%. Contrary to the effect of m_1, Fig. 19.17 demonstrates no sensitivity of predictions of n, r, and p during the initial period of 250 d (6,000 h), and a very short period of no sensitivity for c concentration. Afterward, the temporal patterns of all variables change significantly, with different ranges of concentrations for rods r and cocci c. As in Figs. 19.14 and 19.15, the reference (baseline) time series curves are shown in black color.

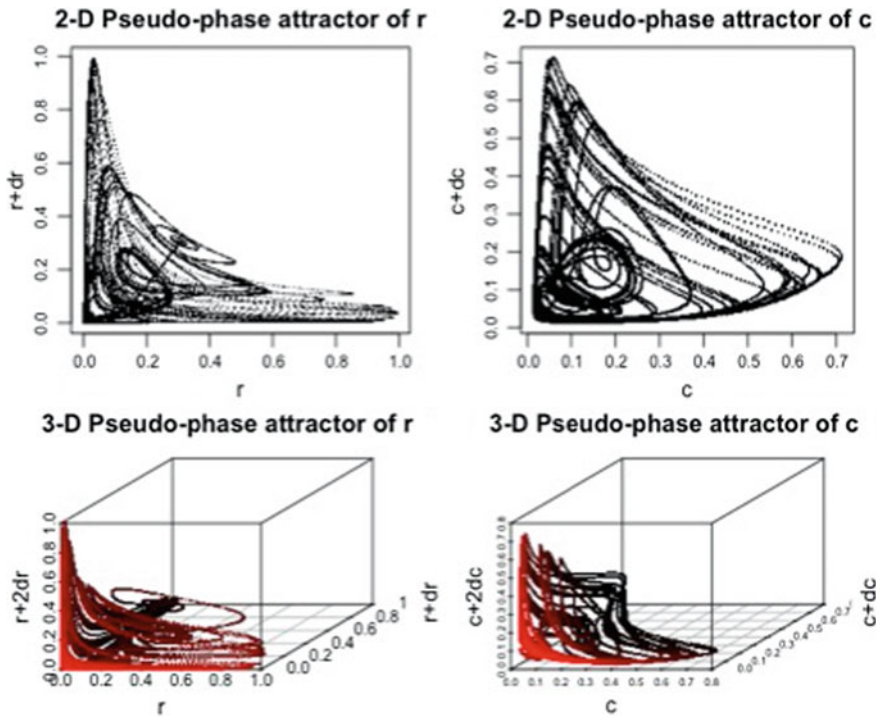

Fig. 19.13 2-D and 3-D pseudo-phase space attractors for n (left) and p (right)

19.7 Summary and Conclusions

We developed a mathematical model of microbial population dynamics, involving the trophic pathways of three microbes (two feeders and one predator) and a nutrient source, to attempt to simulate the results of the experiments conducted in a chemostat and reported by Becks et al. [5–7]. To derive a new model, we used as a baseline a 3-variable (2 microbes and nutrient) equation of Kot et al. [26], and converted it to a 4-variable equation model. The 4 dependent variables of the new model are: $n(t)$, $r(t)$, $c(t)$ and $p(t)$ – nutrient, rods, cocci and predators, respectively, which are the functions of time t. The parameters that were measured in the Becks et al. experiments [5], such as the dilution rates and average mass of the microbes, were utilized in the new model. One of the important features of the model is taking into account a possible preference of p for r versus c, and to make r more competitive for nutrient than c, as well as to recycle some dying p biomass, which was also consistent with the experimental details. Incorporation of these details led to a system of equations including a modified version of Monod kinetics. For further details relating the new model to the biological details of the Becks et al. [5] experiments and the possible application of Shannon's information theory see Molz et al. [30].

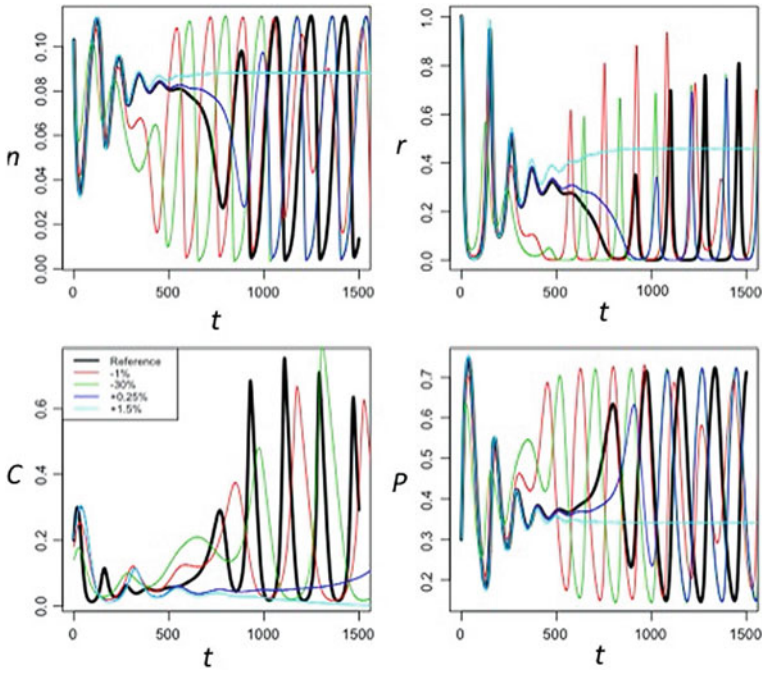

Fig. 19.14 Time series calculations demonstrating the sensitivity of predictions to minor changes in the initial values of C

Numerical simulations were carried out to assess the effects of (a) endogenous factors, i.e., the nutrient flux, $D \cdot n_o$, into the system, (b) exogenous factors, i.e., variations in initial microbial conditions, and (c) model parameters. Nonlinear dynamics analysis of simulated time series was conducted to assess transitions between steady state, deterministic chaotic and quasi-periodic modulations of the nutrient and microbial concentrations. A modeling machine learning framework (modified from [12]) was developed to conduct simulations and the time series analysis in the time domain and phase space based on the application of the high-level statistical computing languages MATLAB [38] and R [39–45]. Calculations in the time domain included the DMA analysis to assess the variations of the Hurst exponent versus the nutrient flux into the system, calculations of the information measure–Shannon's entropy, and the time delay of temporal oscillations of nutrient and microbe concentrations. The phase space analysis of the simulated time series of concentrations was conducted using the methods of nonlinear dynamics and deterministic chaos, in particular, calculations of the global and local embedding dimensions, correlation dimension, information dimension, and the spectrum of Lyapunov exponents. The time series data were used to plot the phase space attractors to express the dependence between the system's state parameters, i.e., microbe concentrations, and pseudo-phase space attractors, in

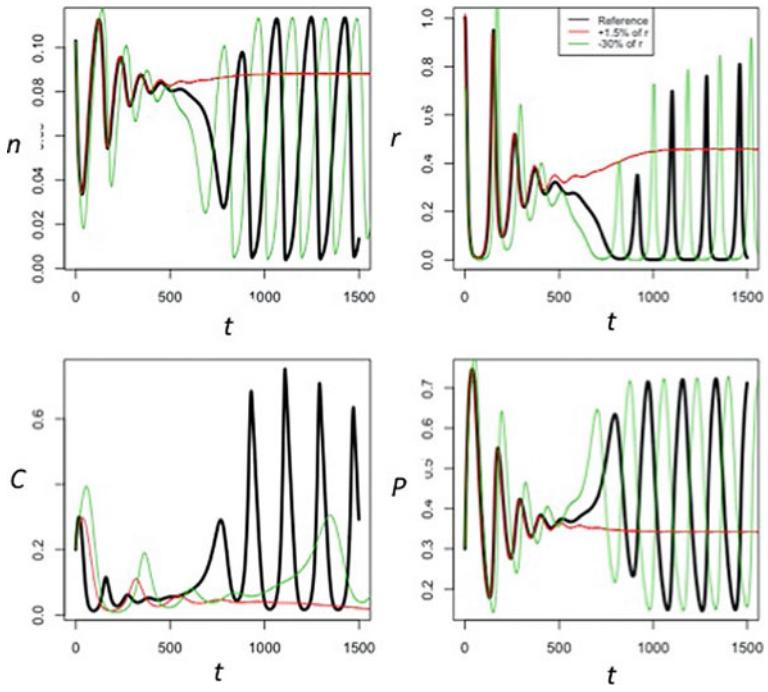

Fig. 19.15 Time series calculations demonstrating the sensitivity of predictions to minor changes in the initial values of *r*

which the attractor axes are used to compare the observations from a single time series, which are separated by the time delay.

Like classical Lorenz or Rossler systems of equations, which generate a deterministic chaotic behavior for a certain range of input parameters, the developed mathematical model generates a deterministic chaotic behavior for a particular set of nutrient fluxes, initial conditions, and model parameters. The simulation results confirmed that even a slight variation of the system's input data might result in vastly different predictions of the temporal oscillations of the system. The simulation results demonstrate a sharp transition from a steady state to deterministic chaotic to quasi-periodic and again to steady state behavior, as the nutrient influx increases. However, as discussed further in Molz et al. [30] it may be possible to further modify the Monod kinetic expressions so that the chaotic domain is enhanced.

Small changes in initial microbial concentrations and input parameters generate extreme changes of temporal chaotic microbial concentrations, although the system behaves within the same attractor. Several types of phase space and pseudo-phase attractors to which a system tends to evolve with time can characterize the simulated system. For small changes in initial conditions, resulting attractors are bounded

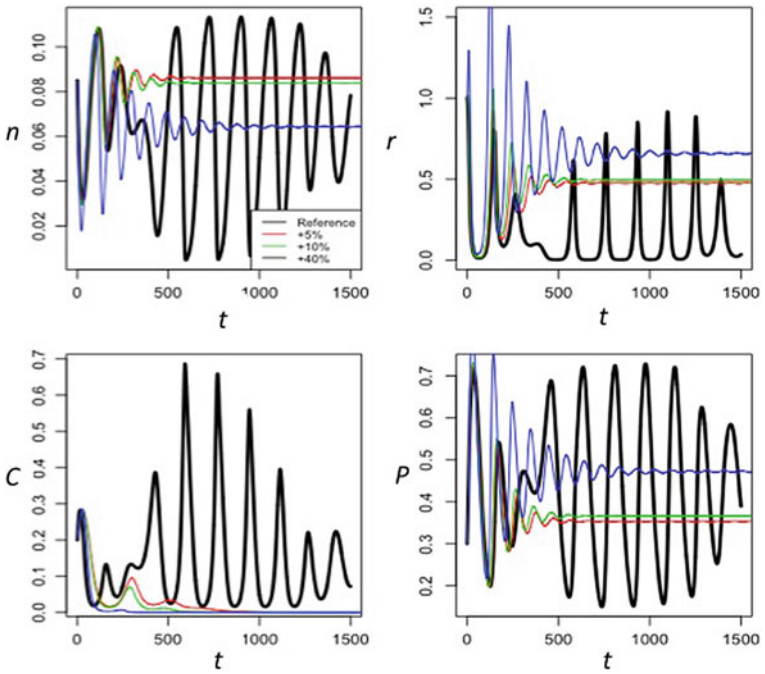

Fig. 19.16 Time series calculations demonstrating the sensitivity of predictions to minor changes in the values of parameter m_1

(contrary to that of a random system), i.e., may represent a 'sustainable state' (i.e., resilience) of the ecological system [27].

This study forms a theoretical basis and reasoning for conducting further experimental studies of nonlinear dynamical processes of microbial systems. Directions of future research may include the application of machine learning algorithms based on both deterministic and stochastic reduced order (surrogate) approximate methods, which can be several orders of magnitude faster than the exact methods for simulations of large spatial-temporal scale ecological processes, while generating practically the same results. We also underline that an integration of experimental results and a mathematical model, both producing classical and deterministic chaotic dynamics, is a useful step in better understanding complex phenomenon in microbial systems. This research is also providing motivation for further study involving spatial variables (i.e., biofilms [14]), and further involvement of Shannon's information theory in biological systems [30].

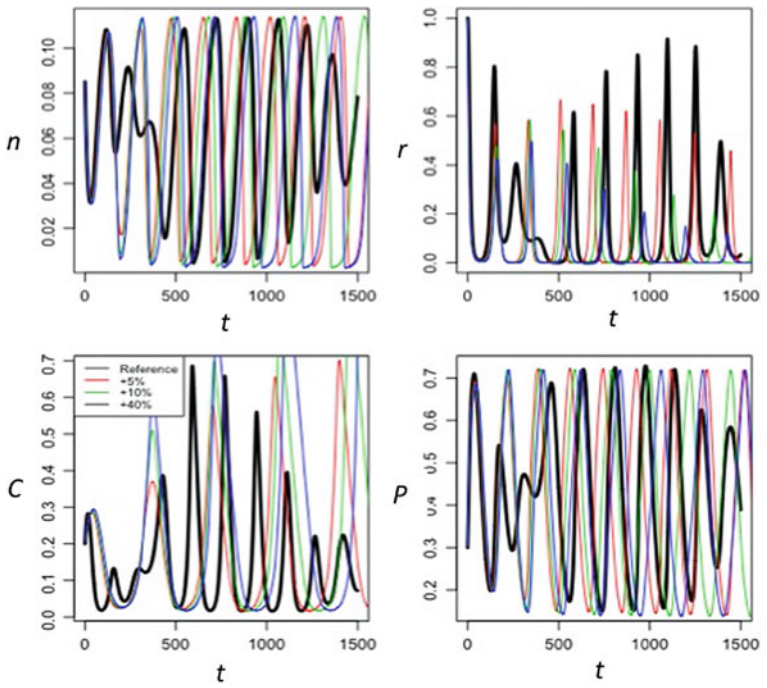

Fig. 19.17 Time series calculations demonstrating the sensitivity of predictions to minor changes in the values of parameter m_2

Acknowledgements BF and DA research supported by the U.S. DOE, Office of Science, Office of Biological and Environmental Research, and Office of Science, Office of Advanced Scientific Computing under the DOE Contract No. DE-AC02-05CH11231. FM acknowledges the support of the Clemson University, Department of Environmental Engineering and Earth Sciences. The invitation by Sergei Silvestrov and Dmitrii Silvestrov to present the paper at the SPAS2017 International Conference on Stochastic Processes and Algebraic Structures, and their kind assistance with LaTeX typesetting of the manuscript for publication are gratefully appreciated.

References

1. Abarbanel, H.D.I.: Analysis of Observed Chaotic Data. Springer, New York (1996)
2. Abarbanel, H.D.I., Tsimring, L.S.: Reader's Guide to Toolkit for Nonlinear Signal Analysis. UCSD, San Diego, California (1998)
3. Abrams, P.A., Ginzburg, L.R.: Coupling in predator-prey dynamics: ratio dependence. Trends Ecol. Evol. **15**(8), 337–341 (1973). https://doi.org/10.1016/s0169-5347(00)01908-x
4. Arditi, R., Ginzburg, L.R.: Coupling in predator-prey dynamics: ratio dependence. J. Theor. Biol. **139**, 311–326 (1989). https://doi.org/10.1016/s0022-5193(89)80211-5
5. Becks, L., Hilker, F., Malchow, H., Jürgens, K., Arndt, H.: Experimental demonstration of chaos in a microbial food web. Nature **435**, 1226–1229 (2005). https://doi.org/10.1038/nature03627

6. Becks, L., Arndt, H.: Different types of synchrony in chaotic and cyclic communities. Nat. Commun. **4** 1359, 9 (2013). https://www.nature.com/articles/ncomms2355
7. Becks, L., Arndt, H.: Transitions from stable equilibria and back in an experimental food web. Ecology **89**, 3222–3226 (2008)
8. Beninca, E., Huisman, J., Heerkloss, R., Johnk, K., Branko, P., Van Nes, E., Cheffer, M., Ellner, S.: Chaos in a long-term experiment with a plankton community. Nature **451**, 822–825 (2008). https://doi.org/10.1038/NATURE06512
9. Curriero, F., Patz, J.A., Rose, J.B., Lele, S.: The association between extreme precipitation and waterborne disease outbreaks in the United States. Am. J. Public Health **91**, 1194–1199 (2001)
10. Codeco, C.T., Lele, S., Pascual, M., Bouma, M., Ko, A.I.: A stochastic model for ecological systems with strong nonlinear response to environmental drivers: application to two water-borne diseases. J. R. Soc. Interface **5**, 247–252 (2008)
11. Dormand, J., Prince, P.: A family of embedded Runke-Kutta formulae. J. Comput. Appl. Math. **6**, 19–26 (1980)
12. Faybishenko, B.: Chaotic dynamics in flow through unsaturated fractured media. Adv. Water Resour. **25**(7), 793–816 (2002)
13. Faybishenko, B., Molz, F.: Nonlinear rhizosphere dynamics yields synchronized oscillations of microbial populations, carbon and oxygen concentrations induced by root exudation. Proc. Environ. Sci. **19**, 369–378 (2013)
14. Faybishenko, B., Murdoch, L., Molz, F.: The Emergence of Chaotic Dynamics in Complex Microbial Systems: Fully Mixed versus 1-D Models. Paper Presented at the Goldschmidt Conference, June 8-13 2014, Sacramento (2014)
15. Hanemaaijer, M., Röling, W., Olivier, B., Khandelwal, R., Teusink, B., Bruggeman, F.: Systems modeling approaches for microbial community studies: from metagenomics to inference of the community structure. Front. Microbiol. **6** (2015). https://doi.org/10.3389/fmicb.2015.00213.
16. Gillman, M., Hails, R.: An Introduction to Ecological Modelling: Putting Practice into Theory. Blackwell Scientific Publications, Oxford (1997)
17. Graham, D., Knapp, C., Van Vleck, E., Bloor, K., Lane, T., Graham, C.: Experimental demonstration of chaotic instability in biological nitrification. ISME J. **1**, 385–393 (2007)
18. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. Physica D**9**, 189–208 (1983)
19. Groffman, P.M., Law, N.L., Belt, K.T., Band, L.E., Fisher, G.T.: Nitrogen fluxes and retention in urban watershed ecosystems. Ecosystems **7**, 393–403 (2004)
20. Grover, J.P.: Resource Competition. Springer, New York (1997)
21. Gunderson, L.: Ecological resilience-theory to practice. Annu. Rev. Ecol. Syst. **31**, 421–439 (2000)
22. Hegger, R., Kantz, H., Schreiber, T.: Practical implementation of nonlinear time series methods: the TISEAN package. Chaos **10**, 413–421 (1999)
23. Hernandez, M.-J.: Dynamics of transitions between population interactions: a nonlinear interaction-function defined. Proc. R. Soc. Lond. B **265**, 1433–1440 (1998)
24. Hirsch, M.: Systems of differential equations which are competitive or cooperative: i. Limit sets. SIAM J. Math. Anal. **13**, 167–179 (1982)
25. Holling, C.S.: Surprise for science, resilience for ecosystems and incentives for people. Ecol. Appl. **6**, 733–735 (1996)
26. Kot, M., Sayler, G., Schultz, T.: Complex dynamics in a model microbial system. Bull. Math. Biol. **54**, 619–648 (1992)
27. Ludwig, D., Walker, B., Holling, C.: Sustainability, stability, and resilience. Conserv. Ecol. **1**, 7, (1997). http://www.consecol.org/vol1/iss1/art7
28. May, R.M.: Thresholds and breakpoints in ecosystems with a multiplicity of stable states. Nature **269**, 471–477 (1977)
29. May, R.M.: Models for two interacting populations. In: May, R.M. (ed.) Theoretical Ecology: Principles and Applications, 2nd edn., pp. 78–104. Sunderland, Sinauer (1981)
30. Molz, F., Faybishenko, B., Agarwal, D.: A Broad Exploration of Nonlinear Dynamics in Microbial Systems Motivated by Chemostat Experiments Producing Deterministic Chaos. LBNL-2001172, Berkeley (2018)

31. Molz, F., Faybishenko, B.: Increasing evidence for chaotic dynamics in the soil-plant-atmosphere system: a motivation for future research. Proc. Environ. Sci. **19**, 681–690 (2013)
32. Rodriguez-Iturbe, I., Entekhabi, D., Bras, R.L.: Nonlinear dynamics of soil moisture at climate scales: 1. Stochastic analysis. Water Resour. Res. **27**(8), 1899–1906 (1991)
33. Rosenstein, M., Collins, J., De Luca, C.: A practical method for calculating largest Lyapunov exponent from small data sets. Physica D **65**, 117–124 (1993)
34. Rosenzweig, M., MacArthur, R.: Graphical representation and stability conditions of predator-prey interactions. Am. Nat. **97**(895), 209–223 (1963)
35. Smale, S.: On the differential equations of species in competition. J. Math. Biol. **3**, 5–7 (1976)
36. Tsonis, A.A.: Chaos: From Theory to Applications. Plenum Press, New York (1992)
37. Tsonis, A., Elsner, J.: Chaos, strange attractors and weather. Bull. Am. Meteorol. Soc. **7**, 14–23 (1989)
38. MATLAB, Release.: The MathWorks Inc. Natick, Massachusetts (2012)
39. R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2016). https://www.R-project.org/
40. Constantine, W., Percival, D.: Fractal: A Fractal Time Series Modeling and Analysis Package. R Package Version 2.0-4 (2017). https://CRAN.R-project.org/package=fractal
41. Hausser, J., Strimmer, K.: Entropy: Estimation of Entropy, Mutual Information and Related Quantities. R package version 1.2.1 (2014). https://CRAN.R-project.org/package=entropy
42. Garcia, C.A.: NonlinearTseries: Nonlinear Time Series Analysis. R package version 0.2.3 (2015). https://CRAN.R-project.org/package=nonlinearTseries
43. Di Narzo, A.F.: TseriesChaos: Analysis of Nonlinear Time Series. R package version 0.1-13. https://CRAN.R-project.org/package=tseriesChaos
44. Ligges, U., Machler, M.: Scatterplot3d - an R package for visualizing multivariate data. J. Stat. Softw. **8**(11), 1–20 (2003)
45. Zeileis, A., Grothendieck, G.: zoo: S3 infrastructure for regular and irregular time series. J. Stat. Softw. **14**(6), 1–27 (2005). https://doi.org/10.18637/jss.v014.i06

Author Index

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1

Subject Index

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1

Q